

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 5 月 2 7 日
Date of Application:

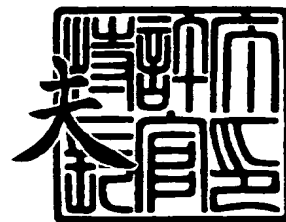
出 願 番 号 特 願 2 0 0 3 - 1 4 8 6 5 7
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 1 4 8 6 5 7]

出 願 人 沖電気工業株式会社
Applicant(s):

2 0 0 3 年 1 1 月 2 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康



【書類名】 特許願

【整理番号】 KN002597

【提出日】 平成15年 5月27日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 G06F 17/28

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

【氏名】 松永 聡彦

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

【氏名】 北村 美穂子

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

【氏名】 村田 稔樹

【特許出願人】

【識別番号】 000000295

【氏名又は名称】 沖電気工業株式会社

【代表者】 篠塚 勝正

【代理人】

【識別番号】 100090620

【弁理士】

【氏名又は名称】 工藤 宣幸

【その他】 国等の委託研究の成果に係る特許出願（平成14年度通信・放送機構「多言語標準文書処理システムの研究開発」に係る委託研究、産業活力再生特別措置法第30条の適用を受けるもの）

【手数料の表示】

【予納台帳番号】 013664

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006358

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書対応関係検査装置、翻訳処理装置、文書対応関係検査方法、翻訳処理方法、および文書対応関係検査プログラム

【特許請求の範囲】

【請求項 1】 第 1 の文書の構成要素と第 2 の文書の構成要素のあいだの対応関係を検査する文書対応関係検査装置であって、

前記第 1 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析すると共に、前記第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析する論理構造解析部と、

当該論理構造解析部の解析結果をもとに、前記第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出する対応関係検出部とを備えたことを特徴とする文書対応関係検査装置。

【請求項 2】 請求項 1 の文書対応関係検査装置において、

前記対応関係検出部は、

同じ文書の文ブロックのなかに階層構造が存在する場合、上位の階層の文ブロックに関する前記対応関係を検出したあと、下位の階層の文ブロックの前記対応関係を検出することを特徴とする文書対応関係検査装置。

【請求項 3】 請求項 1 の文書対応関係検査装置において、

前記対応関係検出部は、

前記第 1 の文書に関する文ブロックと第 2 の文書に関する文ブロックのあいだで所定の類似度を算出する第 1 の類似度算出部を備え、

同じ文書の文ブロックのなかに階層構造が存在する場合、同一の階層の文ブロックのなかで前記類似度が高いものほど優先的に対応関係を検出し、対応関係を検出した文ブロックに対する該当文書上における距離が近い文ブロックほど類似度が高くなるように、前記第 1 の類似度算出部を制御することを特徴とする文書対応関係検査装置。

【請求項 4】 第 1 の文書に関し原文と訳文の対訳を登録した対訳辞書を利用して、当該第 1 の文書の少なくとも一部を変更した改版文書である第 2 の文書

の原文を翻訳処理する翻訳処理装置において、

請求項 1 ～ 3 のいずれかの文書対応関係検査装置と、

前記第 2 の文書に関する原文に含まれる文ブロックのうち、少なくとも、当該文書対応関係検査装置により対応関係が検出された文ブロックに対しては、前記対訳辞書を用いた翻訳処理を実行するブロック翻訳処理部とを備えたことを特徴とする翻訳処理装置。

【請求項 5】 請求項 4 の翻訳処理装置において、

前記文書対応関係検査装置により対応関係が検出された文ブロックの翻訳結果を表示する際、前記第 1 の文書と第 2 の文書の原文間の差分を示す第 1 の差分情報を表示する第 1 の差分情報表示部を備えたことを特徴とする翻訳処理装置。

【請求項 6】 請求項 4 の翻訳処理装置において、

同じ文書の文ブロックのなかに階層構造が存在する場合、前記文書対応関係検査装置により対応関係が検出された文ブロックが属する上位の階層の文ブロックと、前記第 1 の文書の原文との間の差分を示す第 2 の差分情報を表示する第 2 の差分情報表示部を備えたことを特徴とする翻訳処理装置。

【請求項 7】 請求項 4 の翻訳処理装置において、

前記第 1 の文書に関する原文の文ブロックと第 2 の文書に関する原文の文ブロックのあいだで所定の類似度を算出する第 2 の類似度算出部と、

前記文書対応関係検査装置が対応関係を検出した文ブロック以外の文ブロックについても、当該第 2 の類似度算出部が算出した類似度が所定のしきい値以上のものは、対応候補文ブロックとして保存しておき、ユーザとの対話に応じて表示する対応候補処理部とを備えたことを特徴とする翻訳処理装置。

【請求項 8】 第 1 の文書の構成要素と第 2 の文書の構成要素のあいだの対応関係を検査する文書対応関係検査方法であって、

論理構造解析部が、前記第 1 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析すると共に、前記第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析し、

当該論理構造解析部の解析結果をもとに、対応関係検出部が、前記第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出することを特

徴とする文書対応関係検査方法。

【請求項 9】 請求項 8 の文書対応関係検査方法において、
前記対応関係検出部は、

同じ文書の文ブロックのなかに階層構造が存在する場合、上位の階層の文ブロックに関する前記対応関係を検出したあと、下位の階層の文ブロックの前記対応関係を検出することを特徴とする文書対応関係検査方法。

【請求項 1 0】 請求項 8 の文書対応関係検査方法において、
前記対応関係検出部では、

第 1 の類似度算出部が、前記第 1 の文書に関する文ブロックと第 2 の文書に関する文ブロックのあいだで所定の類似度を算出し、

同じ文書の文ブロックのなかに階層構造が存在する場合、同一の階層の文ブロックのなかで前記類似度が高いものほど優先的に対応関係を検出し、対応関係を検出した文ブロックに対する該当文書上における距離が近い文ブロックほど類似度が高くなるように、前記第 1 の類似度算出部を制御することを特徴とする文書対応関係検査方法。

【請求項 1 1】 第 1 の文書に関し原文と訳文の対訳を登録した対訳辞書を利用して、当該第 1 の文書の少なくとも一部を変更した改版文書である第 2 の文書の原文を翻訳処理する翻訳処理方法において、

請求項 8 ～ 1 0 のいずれかの文書対応関係検査方法により前記第 2 の文書に関する原文に含まれる文ブロックと、前記第 1 の文書に関する原文の文ブロックの対応関係を検出し、

前記第 2 の文書に関する原文に含まれる文ブロックのうち、少なくとも、当該文書対応関係検査方法により対応関係が検出された文ブロックに対しては、ブロック翻訳処理部が、前記対訳辞書を用いた翻訳処理を実行することを特徴とする翻訳処理方法。

【請求項 1 2】 請求項 1 1 の翻訳処理方法において、

前記文書対応関係検査方法により対応関係が検出された文ブロックの翻訳結果を表示する際、前記第 1 の文書と第 2 の文書の原文間の差分を示す第 1 の差分情報を表示する第 1 の差分情報表示部を備えたことを特徴とする翻訳処理方法。

【請求項 13】 請求項 11 の翻訳処理方法において、

同じ文書の文ブロックのなかに階層構造が存在する場合、前記文書対応関係検査方法により対応関係が検出された文ブロックが属する上位の階層の文ブロックと、前記第 1 の文書の原文との間の差分を示す第 2 の差分情報を表示する第 2 の差分情報表示部を備えたことを特徴とする翻訳処理方法。

【請求項 14】 請求項 11 の翻訳処理方法において、

第 2 の類似度算出部が、前記第 1 の文書に関する原文の文ブロックと第 2 の文書に関する原文の文ブロックのあいだで所定の類似度を算出し、

前記文書対応関係検査方法が対応関係を検出した文ブロック以外の文ブロックについても、当該第 2 の類似度算出部が算出した類似度が所定のしきい値以上のものは、対応候補処理部が、対応候補文ブロックとして保存しておき、ユーザとの対話に応じて表示することを特徴とする翻訳処理方法。

【請求項 15】 第 1 の文書の構成要素と第 2 の文書の構成要素のあいだの対応関係を検査する文書対応関係検査プログラムであって、コンピュータに、

前記第 1 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析すると共に、前記第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析する論理構造解析機能と、

当該論理構造解析機能の解析結果をもとに、前記第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出する対応関係検出機能とを実現させることを特徴とする文書対応関係検査プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は文書対応関係検査装置、翻訳処理装置、文書対応関係検査方法、翻訳処理方法、および文書対応関係検査プログラムに関し、例えば、ある旧版文書とそれを変更した改版文書（新版文書）のあいだで章、節、文などの対応関係を特定する場合や、このような対応関係の特定結果を用いた翻訳処理を実行する場合などに適用して好適なものである。

【 0 0 0 2 】

【従来の技術】

下記の非特許文献 1 の技術では、あらかじめ翻訳済みの原文、訳文の対訳を「翻訳メモリ」と呼ばれる対訳データベースに格納しておく。翻訳を行うときには、当該対訳データベースを検索した上で、翻訳しようとする原文（対象原文）と比較して最も類似度（一致度）が高い原文を特定する。その類似度がしきい値以上ならば特定した原文と対訳の関係にある訳文を、当該対象原文の翻訳結果として出力する。しきい値以下の場合は何も出力しないか、機械翻訳結果を出力する。

【 0 0 0 3 】

機械翻訳による翻訳結果の品質を高めるには本質的に困難な多くの問題を解決しなければならないが、このように対訳データベースを利用すれば、機械翻訳を行うことなく、高い品質の翻訳結果を得ることができる。

【 0 0 0 4 】

また、複数の翻訳者によりある翻訳プロジェクトを行っている場合、同じ対訳データベースを使用することで用語の訳し方を統一することが可能になる。さらに、例えば、マニュアルや技術文書などのように、改版することが予めわかっている文書の場合、第 1 版の対訳を対訳データベースに格納しておくことで第 2 版以降の改版文書の翻訳作業を効率よく行うことができる。

【 0 0 0 5 】

【非特許文献 1】

『A T L A S V 9 新機能「翻訳メモリ」』（2 0 0 2 年 6 月）

【 0 0 0 6 】

【発明が解決しようとする課題】

ところで、上述した対訳データベースを用いる方法では、文単位の類似度だけを検査し、類似度がしきい値以上であれば、対訳データベースに格納された訳文を翻訳結果として出力しているため、文脈に忠実な翻訳結果を得ることができず、その意味で翻訳の品質が低いといえる。

【 0 0 0 7 】

翻訳処理を行う場合に限らず、適切で詳細な版管理を行う観点から見ても、文単位の類似度だけを検査していたのでは、品質の良い版管理を行うことは難しい。

【0008】

旧版文書に関する対訳を格納した対訳データベースを用いて行う改版文書の翻訳は、より抽象的には、版管理の概念に含まれるものとみることができ、版管理の品質を高めることが翻訳の品質を高めることにもつながる。

【0009】

【課題を解決するための手段】

かかる課題を解決するために、第1の本発明は、第1の文書の構成要素と第2の文書の構成要素のあいだの対応関係を検査する文書対応関係検査装置であって、前記第1の文書の構成要素のうち、少なくとも1つ以上の文を含む文ブロックの論理的な構造を解析すると共に、前記第2の文書の構成要素のうち、少なくとも1つ以上の文を含む文ブロックの論理的な構造を解析する論理構造解析部と、当該論理構造解析部の解析結果をもとに、前記第1の文書の文ブロックと第2の文書の文ブロックのあいだの対応関係を検出する対応関係検出部とを備えたことを特徴とする。

【0010】

また、第2の本発明では、第1の文書に関し原文と訳文の対訳を登録した対訳辞書を利用して、当該第1の文書の少なくとも一部を変更した改版文書である第2の文書の原文を翻訳処理する翻訳処理装置において、請求項1～3のいずれかの文書対応関係検査装置と、前記第2の文書に関する原文に含まれる文ブロックのうち、少なくとも、当該文書対応関係検査装置により対応関係が検出された文ブロックに対しては、前記対訳辞書を用いた翻訳処理を実行するブロック翻訳処理部とを備えたことを特徴とする。

【0011】

さらに、第3の本発明は、1の文書の構成要素と第2の文書の構成要素のあいだの対応関係を検査する文書対応関係検査方法であって、論理構造解析部が、前記第1の文書の構成要素のうち、少なくとも1つ以上の文を含む文ブロックの論

理的な構造を解析すると共に、前記第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析し、当該論理構造解析部の解析結果をもとに、対応関係検出部が、前記第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出することを特徴とする。

【0 0 1 2】

また、第 4 の本発明では、第 1 の文書に関し原文と訳文の対訳を登録した対訳辞書を利用して、当該第 1 の文書の少なくとも一部を変更した改版文書である第 2 の文書の原文を翻訳処理する翻訳処理方法において、請求項 8 ～ 1 0 のいずれかの文書対応関係検査方法により前記第 2 の文書に関する原文に含まれる文ブロックと、前記第 1 の文書に関する原文の文ブロックの対応関係を検出し、前記第 2 の文書に関する原文に含まれる文ブロックのうち、少なくとも、当該文書対応関係検査方法により対応関係が検出された文ブロックに対しては、ブロック翻訳処理部が、前記対訳辞書を用いた翻訳処理を実行することを特徴とする。

【0 0 1 3】

さらにまた、第 5 の本発明では、第 1 の文書の構成要素と第 2 の文書の構成要素のあいだの対応関係を検査する文書対応関係検査プログラムであって、コンピュータに、前記第 1 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析すると共に、前記第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析する論理構造解析機能と、当該論理構造解析機能の解析結果をもとに、前記第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出する対応関係検出機能とを実現させることを特徴とする。

【0 0 1 4】

【発明の実施の形態】

(A) 実施形態

以下、本発明にかかる文書対応関係検査装置、翻訳処理装置、文書対応関係検査方法、翻訳処理方法、および文書対応関係検査プログラムを、翻訳支援システムに提供した場合を例に、実施形態について説明する。

【0 0 1 5】

上述したように、前記対訳データベースを用いる非特許文献 1 の方法では、文単位の類似度だけを検査し、類似度がしきい値以上であれば、対訳データベースに格納された訳文を翻訳結果として出力しているため、文脈に忠実な翻訳結果を得ることができず、その意味で翻訳の品質が低いといえる。

【 0 0 1 6 】

翻訳結果は、1つの文でみた場合には品質が高くても、文と文のつながり、文体、訳調などの観点で統一性に欠けたものであると品質が高いとはいえないからである。また、翻訳結果を得た後でユーザが行う後編集（ポストエディット）の作業効率を高める観点でも、文脈に忠実な翻訳結果であることが望ましい。

【 0 0 1 7 】

例えば、マニュアルなどの旧版に関する対訳を格納した対訳データベースを用いて、当該マニュアルの改版を翻訳する場合、旧版マニュアルと改版マニュアルの文脈に配慮しなければ、改版マニュアルの翻訳結果の品質が低下する可能性が高い。

【 0 0 1 8 】

マニュアルに限らず、例えば自然言語で記述された文書の場合、その文書上における距離（距離は例えば、章、節、段落などの単位で示すことができる。章で示す場合、一例として、同じ章内であれば距離が近く、異なる章に分散していれば距離が遠い）が離れるほど、各種の事情で、用語や言い回しが変化することが多いし、そのような変化は、読者にとって自然なものとして受け入れられる。例えば、まったく同じ表現で記述することも可能な内容を1つの文書のなかで2回（2文）記述する場合、記述される各文の当該文書における距離が近ければそれらの文の表現（用語や言い回し）は同じになることが多いが、距離が遠ければ、用語や言い回しの変化して異なる文となることも少なくない。同様のことは、1つの文書の内部においてのみではなく、文脈に対応関係を持つ傾向の強い2文書間（例えば、同じマニュアルの旧版文書と改版文書のあいだ）でも成立する。

【 0 0 1 9 】

例えば、改版マニュアルの原文のなかに、旧版マニュアルの対訳群のなかのある原文（参照原文）に類似度が高い文（対象原文）が存在する場合、その対象原

文が存在する文脈が、旧版マニュアルにおいて当該参照原文が存在する文脈に対応したものであれば、その参照原文と対訳の関係をなす訳文をそのまま翻訳結果として利用できる可能性が高いが、対応したものでなければ、そのまま翻訳結果として利用できる可能性は低い。また、対応しない文脈であるのに翻訳結果として利用した場合には、後編集によって大幅な変更を行う必要性が高いことが予想されるが、文脈に配慮しない前記非特許文献 1 の技術では、その必要性をユーザに伝える方法がないため、ユーザは結局、高い類似度で得られた訳文に対しても、低い類似度で得られた訳文に対するのと同様な注意深さで後編集作業を行うことを強いられ、後編集の作業効率が低い。

【0 0 2 0】

そこで、本実施形態では、文脈に忠実な翻訳を行うことにより、翻訳結果の品質を高めることを特徴とする。

【0 0 2 1】

(A-1) 第 1 の実施形態の構成

本実施形態にかかる翻訳支援システム 1 0 の全体構成例を図 1 に示す。

【0 0 2 2】

図 1 において、当該翻訳支援システム 1 0 は、入力部 1 と、文書構造解析部 2 と、文書構造比較部 3 と、差分情報生成部 4 と、旧版文書データベース 5 と、制御部 6 と、出力部 7 と、翻訳処理部 8 とを備えている。

【0 0 2 3】

このうち入力部 1 は、例えば、キーボードやマウスなどのポインティングデバイス、スキャナと文字認識処理などの各種機能によって構成され得る部分で、ユーザ U 1 が各種入力操作を行なう際に機能する。

【0 0 2 4】

出力部 7 は、例えば、ディスプレイ装置への表示、音声への変換および音声出力などの各種機能によって構成され得る部分で、ユーザ U 1 に対して各種の情報を提供する。ここで、ユーザ U 1 は、当該翻訳支援システム 1 0 を操作するオペレータなどであってよい。

【0 0 2 5】

ただし当該入力部 1 や出力部 7 は、人間であるユーザ U 1 とのインタフェースとして機能するだけでなく、リモートの、あるいはローカルの情報処理装置（図示せず）とのあいだで制御情報やデータのやり取りを行うためにも機能し得る。このようなユーザ U 1 あるいは情報処理装置とのやり取りに応じて、旧版データベース 5 の格納内容などが増減、変更されるものであってもよい。また、旧版データベース 5 の本体は Web サーバ側などに配置しておき、検索結果のみ（あるいは、翻訳結果のみ）をネットワーク経由で当該翻訳支援システム 10 に取得する構成としてもよい。検索結果のみを取得するには、Web サーバ側で CGI プログラムなどを利用して検索を行い、その結果を翻訳支援システム 10 へ返送するようにすればよい。

【0026】

制御部 6 は、ハードウェア的には当該翻訳支援システム 10 の CPU（中央処理装置）に相当し、ソフトウェア的には OS（オペレーティングシステム）などの各種プログラムに相当する部分である。当該制御部 6 により、翻訳支援システム 10 内の他の構成要素 1～5，7，8 が制御され得る。

【0027】

前記旧版データベース 5 自体は基本的に上述した対訳データベースに相当する構成要素で原文（の 1 文）を指定すればその訳文（の 1 文）が抽出できるように構成されているが、本実施形態では対訳の利用法が前記非特許文献 1 とは相違するため、その相違に応じて、データベースの格納内容にも従来とは異なる点がある。この旧版データベース 5 には、例えば、マニュアル、技術文書、論文などのように、改版することが予想される文書の旧版（例えば、第 1 版）を格納しておく。旧版データベース 5 には同時に複数の旧版文書（例えば、ある機種のパーソナルコンピュータに関するマニュアルの旧版文書と、他の機種のパーソナルコンピュータに関するマニュアルの旧版文書など）を格納しておくことができるが、以下では、旧版データベース 5 内に格納されている 1 つの文書 DC 1 に注目して説明を進める。

【0028】

なお、一般的には、1 つの原文の文章とその翻訳結果である訳文の文章は別個

の文書とされるが、ここでは、当該文書DC1は原文文章(OR1)の内容と訳文文章(CP1)の内容を含む1つの対訳文書であるものとする。

【0029】

原文文章は、ある内容を第1言語(原文言語(例えば、日本語))で表現するために順序づけられた文の集合体であり、訳文文章は、ある内容を第2言語(訳文言語(例えば、英語))で表現するために順序づけられた文の集合体である。一般的には、原文文章中の各文とその翻訳結果である訳文文章の各文とは1対1に対応しないが、当該文書DC1は対訳文書であるため、原文文章OR1中の各文と、訳文文章CP1中の各文は、1対1に対応している。したがって、文脈(文脈は、後述する階層構造にも対応)の観点でも、当該原文文章OR1と訳文文章CP1は厳密に対応している。

【0030】

旧版データベース5の内部は、原文文章OR1を格納する旧版原文データベース5Aと、訳文文書CP1を格納する旧版訳文データベース5Bに分けることができる。

【0031】

前記文書構造解析部2は文書の構造を解析し、その解析結果を文書構造比較部3に供給する部分である。ここで、構造とは、文章の自然言語的かつ論理的な構造のことで、1つの文章上における例えば、章、節、段落、文などの位置や包含、被包含の関係などに関する構造を指す。多くの場合、前記マニュアル、技術文書、論文などのように論理的な構造が比較的明確な文章では、1つの文章のなかに複数の章があり、各章のなかには1または複数の節があり、各節のなかには1または複数の段落があり、各段落のなかには1または複数の文があるというように階層構造を備えている。したがって、当該文書構造解析部2の役割は、当該階層構造を解析することである。

【0032】

ここで、章、節、段落を、少なくとも1つ以上の文の集まりという意味でブロックと呼ぶ。文は、ブロックの概念のなかに含めることもできるが、ここでは、含まないものとしている。当該ブロック間には前記階層構造がある。なお、通常

、1つの節には、1または複数の段落が含まれるが、ここでは、簡単のために段落を無視し、節のブロックに対しては、直接、文が含まれるものとする。

【0033】

当該文書構造解析部2による解析の対象となる文書は、前記入力部1を介して入力される改版文書DC2中の文章である改版文章OR2と、前記旧版文書DC1に含まれる旧版原文文章OR1である。ただし、旧版原文文章OR1のほうは予め内容が決まっているため、改版文章OR2が得られるまえに解析を行い、解析結果を旧版原文データベース5A内に格納しておくことが可能である。この点は旧版訳文文書CP1についても同様である。処理の効率を高めるためには、旧版原文文書OR1や旧版改版文書CP1の階層構造はあらかじめ解析して旧版データベース5などに格納しておくことが望ましい。

【0034】

この旧版原文文章OR1の内容の一例を抽象化して示したものが、図2(A)である。同様に、改版原文文章OR2の内容の一例を抽象化して示したものが図2(B)である。

【0035】

図2(A)および(B)において、アンダーラインを付与した「1」や「2」などは、章の番号である。また、「1. 1」や「2. 2」などは、「.」の左側の数字が章の番号を示し、右側の数字がその章に含まれる節の番号を示す。したがって、例えば、「1. 1」は第1章第1節を示す。

【0036】

図2(A)において、「文1」、「文2」、「文5」などは各節に含まれる文を示している。ここで、「文」のあとに記述する数字(文識別子)の異同によって、その文の内容をなす文字列の異同を表現している。したがって、「文1」と「文2」は異なる文である。また、図2(A)において、例えば、第1章第2節と第4章には、ともに、「文6」で示された同じ文が存在する。

【0037】

改版原文文章OR2を示す図2(B)も基本的にこの図2(A)と同じである。同じ文書(例えば、同じ機種のパersonalコンピュータに関するマニュアル)

の旧版と改版の関係にあるため、これら 2 つの文章 OR 1、OR 2 には内容的に共通する部分が多い。

【0038】

ただし図 2 (B) では、「文 A」、「文 B」などと、文識別子として数字ではなく、アルファベットを使用している。「文 A (1)」、「文 B (2)」などと、括弧内に記述した数字は、図 2 (A) に示す原文文章 OR 1 側の文識別子で、旧版と改版のあいだで文の対応関係を示している。

【0039】

なお、本実施形態で、文を識別するための識別情報としては当該文識別子のほかに文番号がある。文識別子は文の内容をなす文字列を識別する情報であるのに対し、文番号のほうは、文が当該原文文章上に出現した順番を示す情報である。

【0040】

このように文番号は、各原文文章ごとに、文章上に出現した順番（図 2 (A) および (B) の上から順番）に付与されるため、同じ文字列の文（同じ文識別子を付与された文）であっても、文章上の位置が変われば、文番号も変わる。したがって、図 2 (A) において第 1 章第 2 節と第 4 章に出現する上述した「文 6」には、第 1 章第 2 節で出現したときと第 4 章に出現したときでは、それぞれ別な文番号が付与される。

【0041】

図 2 (A) に示す旧版原文文章 OR 1 の文と文番号の対応関係は、図 15 に示す文・文番号対応表の通りである。また、文番号をもとに、旧版原文文章 OR 1 と改版原文文章 OR 2 の各文の対応関係をまとめると、図 16 に示す新旧文対応表が得られる。

【0042】

改版文書 DC 2 や旧版文書 DC 1 が、その論理的な構造が所定の定型的方法で明示された文書（例えば、HTML 文書や XML 文書などのマークアップ言語で記述された文書）であることは、文書構造解析部 2 が行う解析処理を簡略化する上で望ましいが、必ずしもその必要はない。

【0043】

図 2 (A) および (B) の文章を前提とすると、文書構造解析部 2 による解析結果は、図 4 (A) および (B) の構造情報表の形に整理することができる。図 4 (A) は、旧版原文文章 O R 1 に関する解析結果を整理したものであり、図 4 (B) は、改版原文文章 O R 2 に関する解析結果を整理したものである。

【 0 0 4 4 】

図 4 (A) および (B) において、ブロック番号は、上述したブロックが各原文文章上に出現した順番に付与した番号である。また、階層位置とは、階層の深さのことである。階層構造は木構造で表現できるが、深さ 0 がその文章全体（例えば、旧版原文文章 O R 1 全体または改版原文文章 O R 2 全体）に対応する木の根（ルート）であるとする、深さ 1 は前記章に対応する木のノードであり、深さ 2 は前記節に対応する木のノードである。深さ 3 は後述する副次節に対応するノードであり、改版原文文章 O R 2 のほうにだけ存在する。

【 0 0 4 5 】

下位ブロック番号は、各ブロックより深さが 1 つだけ深く、各ブロックに属するブロックのブロック番号である。また、文番号は、該当するブロック番号で指定されるブロックに属する文の文番号である。

【 0 0 4 6 】

さらに、対応ブロック番号及び類似度は、旧版原文文章 O R 1 と改版原文文章 O R 2 のあいだで対応関係が確定できたブロックのブロック番号と、その確定の根拠となった類似度である。類似度の詳細については後述するが、図示の状態ではまだ対応関係が確定できたブロックが存在しないため、対応ブロック番号及び類似度は、空欄となっている。

【 0 0 4 7 】

対応ブロック番号及び類似度の内容としては、相互に対応するもの（対称的な内容）が記述されることになるので、データ項目としての「対応ブロック番号及び類似度」は、必ずしも図 4 (A) と (B) の双方に設ける必要はなく、例えば、図 4 (B) にのみ設けるようにしてもよい。

【 0 0 4 8 】

前記文書構造比較部 3 は、当該文書構造解析部 2 の解析結果である階層構造を

利用して、前記改版原文文章 O R 2 と旧版原文文章 O R 1 の論理的な構造を比較する部分である。両者の論理的な構造を比較すれば、文のレベルで対応していることが確認された改版原文文章 O R 2 のブロックに関しては、その訳文として、旧版訳文文章 C P 1 の該当ブロックの内容をそのまま利用でき、対訳による翻訳が可能になる等の利点がある。

【 0 0 4 9 】

この比較を行うため、当該文書構造比較部 3 は、階層照合部 3 A と、詳細照合部 3 B とを備えている。

【 0 0 5 0 】

階層照合部 3 A は、前記改版原文文章 O R 2 と旧版原文文章 O R 1 のあいだで、階層構造の深さを比較する部分である。改版によって文章の階層構造の深さが変化し、例えば、図 2 (B) の「3. 2」における「3. 2. 1」、「3. 2. 2」のように、節と文のあいだに新たな階層（副次節）が設けられること等があり得るが、前記詳細照合部 3 B の処理を行うためには、階層構造の深さを揃えておく必要があるため、当該階層照合部 3 A が必要になる。したがって、詳細照合部 3 B が行う処理の具体的な仕様によっては、当該階層照合部 3 A を省略できる可能性もある。

【 0 0 5 1 】

詳細照合部 3 B は、旧版原文文章 O R 1 と改版原文文章 O R 2 のあいだで、各ブロックの対応関係を検査する部分である。この検査（すなわち、ブロック対応決定処理）のため、当該詳細照合部 3 B は、旧版原文文章 O R 1 と改版原文文章 O R 2 のあいだで各文の異同（各文の文字列の異同）を検査する。また、当該詳細照合部 3 B は、ブロックが対応するか否かを識別する際の基準となるしきい値 T H 1 の設定を受けている。後述するように類似度の最大値が 1 0 0 % で最小値が 0 % の場合、しきい値 T H 1 は 1 0 0 % と 0 % の中間の値に設定されることになる。しきい値 T H 1 の値はどのように決めてもよいが、一例としては、4 0 % に設定してもよい。

【 0 0 5 2 】

ブロックの対応関係は、同じ階層位置における文章 O R 1 と O R 2 のすべての

ブロックの組み合わせの類似度を計算した上で、当該類似度に基づいて決定する。

【0053】

また、類似度を求めるのは、改版原文文章OR2上のあるブロック（すなわち、木のノード）と対応する旧版原文文章OR1上の1つのブロックを探索するためであるから、この組み合わせが、1対のブロックから構成される組み合わせであるのは当然である。

【0054】

類似度は、1対のブロックの類似の度合いを示すことができる方法であれば、どのような計算方法によって算出してもかまわないが、次の式（1）にしたがって算出することは簡便である。

【0055】

$$100 \times \text{完全一致する文数} / ((1 \text{ 対のブロックの総文数}) / 2) \cdots (1)$$

図2（A）および（B）において、階層位置2について調べる場合、例えば、1対のブロックとして旧版原文文章OR1の第1章第1節と改版原文文章OR2の第1章第1節の組み合わせを選ぶときには、式（1）中の1対のブロックの総文数は、8（＝4＋4）となり、完全一致する文数は、4となるから、類似度は100%である。

【0056】

同様に、1対のブロックとして旧版原文文章OR1の第1章第2節と改版原文文章OR2の第1章第1節の組み合わせを選ぶときには、式（1）中の1対のブロックの総文数は、7（＝3＋4）となり、完全一致する文数は、0となるから、類似度は0%である。これと同様な検査を、階層位置2のブロックに関するすべての組み合わせについて実行する。異なる階層位置についても同様である。

【0057】

なお、式（1）では、同一ブロック内における変化に関する限り文の出現位置の変化（相対的な出現位置の変化）は反映されないが、改版では、文の文字列は変化していなくても文が出現する位置が替わることがあるので、そのような位置の変化も、類似度に反映されるようにすることは望ましい。

【0058】

図4 (A) および (B) のケースについて、例えば、階層位置2のブロックに関する組み合わせを、(文章OR1内のブロックのブロック番号, 文章OR2内のブロックのブロック番号)の形式にしたがって列挙すると、(2, 2)、(2, 3)、(2, 6)、(2, 7)、(3, 2)、(3, 3)、(3, 6)、(3, 7)、(5, 2)、…、(10, 6)、(10, 7)となる。

【0059】

改版によって、旧版(例えば、OR1)には存在しなかった新たな章や節が改版文章(例えば、OR2)に出現したり、章や節の内容が部分的に変更されること等が起り得るが、新たに出現した章や節の場合、前記詳細照合部3Bは、旧版原文文章側に対応するブロックは存在しない旨の決定を行う。また、改版によって、章や節の内容が部分的に変更された場合には、旧版原文文章側に対応するブロックは存在するものの、その類似度は低いことを示す。

【0060】

単純に前記式(1)にしたがって各組み合わせの類似度を求めることによって、各ブロックの対応関係を決定(対応するブロックが存在しない旨の決定なども含む)することが可能であるが、本実施形態の詳細照合部3Bは、浅い階層位置から順番に前記類似度を求め、深い階層位置で類似度を求めるときには、式(1)で得られた結果をそのまま用いずに、深い階層位置のブロックが属する浅い階層位置のブロック(深い階層位置のブロックからみると、この浅い階層位置のブロックは、親ブロック(上位ブロック)にあたる)の対応関係の検査結果に応じて変更する。

【0061】

この変更は、対応関係が決定できたブロック(対応確定ブロック)に属するブロックに比べ、対応するブロックが存在しない旨の決定を行ったブロック(非対応確定ブロック)に属するブロックの類似度は低くなるように制御することによって実現する。この制御は、例えば、所定の係数 ρ ($0 < \rho < 1$)を、式(1)で算出した類似度に乗算するものであってよい。また、 ρ の具体値は、例えば、0.8や、0.9であってよい。係数 ρ の値は、1つだけとしてもよいが、複数

を用意することも望ましい。

【 0 0 6 2 】

係数 ρ を複数用意する場合、対応確定ブロックに属するブロック（このブロックからみると、当該対応関係確定ブロックは、親ブロック（上位ブロック）にあたる。反対に、親ブロックである当該対応確定ブロックからみると、対応確定ブロックに属するブロックは、子ブロックにあたる）であっても、その対応確定ブロックの対応関係決定の根拠となった類似度の値の高低に応じて、 ρ の値を変化させる。これは、根拠となった類似度の値が小さいほど、乗算する係数 ρ の値も小さくし、式（１）によって算出された類似度を小さいほうへ変化させるものである。

【 0 0 6 3 】

これにより、原文文章 O R 1 と O R 2 のあいだの親ブロック間の対応関係により、子ブロックの対応関係が規制されるようになるから、親ブロックの範囲を越えて子ブロックの対応関係が確定する可能性を、確率的に低減することができる。これは、改版により文の一部を変更し、旧版と改版のあいだでその文自体の類似度は低くなってしまったようなケースでも、全体の文脈がそれほど大きく変化していなければ、旧版と改版のあいだでその文を対応づけることができることを意味する。前記非特許文献 1 の技術の場合、このようなケースでは、対訳による翻訳を行うことはできないが、本実施形態では、このようなケースにおいても、対訳による翻訳を行うことができる。

【 0 0 6 4 】

もちろん、当該文に注目する限り、その翻訳結果は正しいものではないが、それは、後編集によって効率的に修正することができる。

【 0 0 6 5 】

前記翻訳処理部 8 は、当該文書構造比較部 3 の処理を受けて、改版原文文章 O R 2 の翻訳処理を実行する部分で、当該翻訳処理に応じて、改版原文文章 O R 2 の訳文にあたる改版訳文文章 C P 2 を出力する。

【 0 0 6 6 】

本実施形態において、改版原文文章 O R 2 の翻訳は主として、改版原文文章 O

R 2 中のブロックを、対応する旧版訳文文章 C P 1 中のブロックで置き換えることによって実行する。旧版原文文章 O R 1 と旧版訳文文章 C P 1 は厳密に対応しているため、改版原文文章 O R 2 上の対応確定ブロックであれば必ず、旧版訳文文章 C P 1 中に対応するブロックを有する。この場合のブロックは、できるだけ下位の階層のブロック（例えば、節のブロック）とすることが望ましい。

【0067】

なお、改版原文文章 O R 2 上の非対応確定ブロックの場合、旧版訳文文章 C P 1 中に対応するブロックを持たないため、このようなブロック単位の置き換えによる翻訳を行うことができない。したがって、改版原文文章 O R 2 上の非対応確定ブロックの翻訳にあたっては、例えば、通常の機械翻訳を用いたり、前記非特許文献 1 で行われているように、文の類似度に基づいて、（ブロック単位ではなく）文単位で、旧版データベース 5 を利用する対訳による翻訳を行うようにしてもよい。

【0068】

通常の機械翻訳では、形態素解析、構文解析など周知の各種処理の処理結果を利用して、動的に翻訳処理を実行する。

【0069】

前記類似度が 100% ではないブロックであっても、可能な限り機械翻訳を行わず、対訳による翻訳を行うことによって、後編集の作業効率を高めることができる。対訳による翻訳のほうが、機械翻訳による翻訳よりも、文と文のつながり、文体、訳調などの統一性に優れているからである。

【0070】

前記差分情報部 4 は、旧版訳文文章 C P 1 と改版訳文文章 C P 2 の差分に相当する情報（補助情報）を出力する部分である。この補助情報により、例えば、前記ディスプレイ装置の表示画面上で、改版によって削除された旧版原文文章 O R 1 や旧版訳文文章 C P 1 中のブロックを指定できるほか、改版訳文文章 C P 2 中の機械翻訳を行ったブロックを指定するために利用することもできる。機械翻訳を行ったブロックは、後編集の必要性が高いブロックである。改版訳文文章 C P 2 が長い文章であるとしても、画面上で当該補助情報を目視したユーザ U 1 は、

当該補助情報によって指定されるブロックのみに注力して後編集を行うことができるから、後編集の効率が高まる。

【0071】

なお、前記旧版データベース5が、ハードディスクや光ディスクなどの不揮発性記憶手段や、メモリなどの揮発性記憶手段などの記憶資源の上に構築されることは当然である。

【0072】

以下、上記のような構成を有する本実施形態の動作について、図3、図5、図6のフローチャートを参照しながら説明する。

【0073】

図3と図5のフローチャートは一連の全体的な処理の流れを示すもので、図3のフローチャートの処理につづいて図5のフローチャートの処理が実行される。図3のフローチャートは、S10～S14の各ステップから構成され、図5のフローチャートは、S15～S27の各ステップから構成されている。

【0074】

また、図6のフローチャートは、前記詳細照合部3Bが行う各ブロックの対応関係の検査（ブロック対応決定処理）の詳細を示すフローチャートであり、S30～S36の各ステップから構成されている。図5との関係では、この図6のフローチャートは、図5中のステップS19、S22、またはS26の詳細動作を示すものである。

【0075】

上述した説明から明らかなように、これら図3、図5、図6のフローチャートは、旧版原文文章OR1と改版原文文章OR2に関して実行される処理である。

【0076】

（A-2）第1の実施形態の動作

図3において、マニュアルなどの旧版文書DC1に含まれる旧版原文文章OR1と旧版訳文文章CP1が旧版データベース5に格納された状態で、前記改版（新版）原文文章OR2を内容とする改版文書DC2が入力部1から供給されたものとする。この供給は、翻訳支援システム10に対し、当該改版原文文章OR2

の翻訳を要求する指令をとまう。

【0077】

本実施形態において文章OR1とOR2を当該翻訳支援システム10で処理するためには、2つの文書が、前記文書構造解析部2によって解析され、図4（A）および（B）に示す構造情報表の形式で整理されている必要がある。上述したように、旧版原文文章OR1があらかじめ解析され、その階層構造が得られている場合にはその必要はないが、そうでない場合には、図4（A）の構造情報表を得るために解析を行うことになる（S10、S11）。このとき、図15の文・文番号対応表も得られる。

【0078】

改版原文文章OR2に対しても同様な解析が行われ図4（B）の構造情報表が得られる（S12）。

【0079】

次に、文章OR1とOR2のうち階層構造の深さが浅いほうの最も深い階層位置の値を最大階層数を示す最大階層変数Max Layerに代入する。これは、2つの文章OR1、OR2の階層構造の深さを浅いほうに揃えるためである。同時に、階層構造表の不要ブロックレベル行を削除する（S13）。この削除は、2つの文章OR1、OR2で深さが揃っていないときに行う処理である。図2（A）および（B）の例では、この削除により、図2（B）の「3．2．1」と「3．2．2」に対応する図4（B）上の2つの行が削除され、最大階層変数Max Layerには、2が代入される。

【0080】

次に、図15に示す文・文番号対応表を用いて、改版原文文章OR2中の各文に完全一致する旧版原文文章OR1中の文を調べて、図16に示す上述した新旧文対応表を作成する（S14）。

【0081】

このステップS14につづき、図5のステップS15では、検査階層変数iに1を代入する。この変数iは、これからブロック間の対応関係を検査する階層位置を示す変数である。上述したように、ブロック番号自体には階層位置の相違は

反映されていないため、検査階層変数 i によって、前記詳細照合部 3 B が行うブロック対応決定処理の対象とする階層位置を制御する必要がある。換言するならば、階層位置の相違を反映したブロック番号を付与するようにすれば、図 5 のフローチャートの内容は大きく異なるものになる可能性がある。

【0082】

当該ステップ S 1 5 で、検査階層変数 i に初期値である 1 を代入することにより、階層位置 1、すなわち、上述した章のレベルにおけるブロック間の対応関係の検査（ブロック対応決定処理）が開始される。上述したように、階層位置としては 0 もあるが、ここでの初期値は 1 である。

【0083】

この階層位置 i におけるブロックに関し、すべての前記組み合わせを処理するため、ブロック対応決定処理を行っていないブロック（このブロックのブロック番号を j とする）の選択（S 1 6）と、下位ブロックのブロック番号が当該 j である上位ブロック（このブロックのブロック番号を k とする）の選択（S 1 7）を行う。

【0084】

次に、ブロック番号が当該 k である上位ブロックと対応するブロック（このブロックのブロック番号を m とする）が、旧版原文文章 O R 1 側に存在するか否かを検査し（S 1 8）、存在する場合には、ブロック番号が k 、 m それぞれの当該上位ブロックを親ブロックとする下位ブロック（子ブロック）をすべて選択して、それらに対してブロック対応決定処理を行い（S 1 9）、存在しない場合には、処理をステップ S 2 0 へ進める。

【0085】

ただし、階層位置が 1 の場合、その上位ブロック（親ブロック）は階層位置 0、すなわち、その原文文章全体のブロックしかない。文書 D C 1 と D C 2 は、例えば、ある機種のパーソナルコンピュータに関するマニュアルなど、同一の文書の旧版と改版の関係にあるから、当然、階層位置 i が 1 の回の処理では、前記ステップ S 1 8 は無条件に Y E S 側に分岐させることになる。

【0086】

ステップ S 2 0 では、改版原文文章 O R 2 内の階層位置 i のブロックに対する全ての上位ブロック（全ての親ブロック）について、ブロック対応決定処理を行ったか否かを調べ、ブロック対応決定処理を行っていない親ブロックがある場合には、前記ステップ S 1 6 に戻って同様な処理を繰り返す。すべての親ブロックについてブロック対応決定処理が終わった場合には、処理はステップ S 2 1 へ進む。ステップ S 2 1 では、図 4（B）の構造情報表の該当する行（該当するブロック）のうち、対応ブロック番号及び類似度の欄が未記入のものがあるか否かを調べる。この欄が未記入の行は、前記ブロック対応決定処理を終えていないブロック（対応未決定（対応未確定）のブロック）の行であるため、前記ブロック対応決定処理を行うことになる（S 2 2）。

【0 0 8 7】

当該階層位置 i のすべてのブロックにつき、対応関係（対応確定ブロックまたは非対応確定ブロックの別）が確定すると、その時点の i の値が前記最大階層変数 $Max Layer$ の値より小さいか否かを検査する（S 2 3）。小さければステップ S 2 3 は Y E S 側に分岐して、 i をインクリメントし（S 2 4）、処理を前記ステップ S 1 6 に戻すことになるが、小さくなければステップ S 2 3 は N O 側に分岐して処理はステップ S 2 5 へ進む。ここでは、 $Max Layer$ の値は 2 であるから、 i の値が 1 であると、ステップ S 2 3 は Y E S 側に分岐することになる。

【0 0 8 8】

ステップ S 2 5 では前記ステップ S 2 1 と同様に、対応ブロック番号及び類似度の欄が未記入のものがあるか否かを調べ、あればそのブロックについても前記ブロック対応決定処理を実行する。ステップ S 2 3 が N O 側に分岐したあとで実行されるのであるから、当該ステップ S 2 6 では、階層位置がもっとも深い 2 のブロック（すなわち、節）について対応関係が決定され、改版原文文章 O R 2 に含まれるすべてのブロックに対する対応関係が確定することになる。

【0 0 8 9】

もちろん、この確定により、いずれのブロックとも対応しない（対応するブロックが存在しない）前記非対応確定ブロックが出現し得ることは当然である。

【0090】

次に、前記ステップS19, S22, S26の詳細動作にあたるブロック対応決定処理の詳細を、図6のフローチャートに基づいて説明する。

【0091】

図6において、すでに処理を行う階層位置などは決まっているので、その階層位置におけるすべてのブロックの組み合わせを求め、各組み合わせにつき、前記式(1)に応じた類似度を算出した上で、類似度の大きい順番に整列して図17に示すブロック組み合わせ表を作成する(S30)。類似度は単純に式(1)にしたがって算出するだけでなく、上述した係数 ρ を乗算してもよいことはすでに説明した通りである。

【0092】

図1.7は、図4(A)および(B)の構造情報表を前提とした階層位置が1の場合のブロック組み合わせ表である。図1.8からも明らかなように、図4(A)の階層位置1には、ブロック番号1, 4, 8, 11の各ブロックが存在し、図4(B)の階層位置1には、ブロック番号1, 4, 5, 10の各ブロックが存在する。同様の関係は、図1.9(A)および(B)にも図示している。図1.9(A)から明らかなように、例えば、改版原文文章OR2のブロック番号1のブロック(章)にはブロック番号2と3の各ブロック(節)が属し、ブロック番号5のブロックにはブロック番号6と7の各ブロックが属している。同様に、図1.9(B)で、旧版原文文章OR1のブロック番号1のブロック(章)にはブロック番号2と3の各ブロック(節)が属し、ブロック番号4のブロックにはブロック番号5, 6, 7の各ブロックが属している。

【0093】

図1.7に示すブロック組み合わせ表の内容を、上述した(文章OR1内のブロックのブロック番号, 文章OR2内のブロックのブロック番号)の形式にしたがって表記すると、ステップS30で作成するブロックの組み合わせの最も上の行L21は、(8, 10)となり、上から2番目以降の行L22~L26は順次、(1, 1)、(4, 5)、(11, 1)、(4, 4)、(4, 1)となる。

【0094】

次に、当該ブロック組み合わせ表の各行のなかから、類似度のもっとも高い組み合わせに対応する行（ここでは、L 2 1）を選択し（S 3 1）、その行の類似度の値が所定のしきい値 T H 1 以上であるか否かを検査する（S 3 2）。

【0095】

類似度のもっとも高い組み合わせでさえ、しきい値 T H 1 未満であれば、対応関係のあるブロックは存在しないことを意味するから、前記対応確定ブロックは得られず、前記非対応確定ブロックのみが得られて処理が終わる。

【0096】

ただし同じ文書の旧版と改版の関係にある文書 D C 1 と D C 2 の場合、すべての組み合わせの類似度がしきい値 T H 1 未満となることは、ほとんどあり得ないので、多くの場合、いくつかの組み合わせでは、類似度がしきい値 T H 1 以上となり、対応確定ブロックが得られる。したがって、多くの場合、類似度が最大の組み合わせである行 L 2 1 については、対応確定ブロックが得られる。

【0097】

しきい値 T H 1 を前記 40% に設定したものとすると、図 17 の例では、行 L 2 1 ～ L 2 4 の組み合わせでは、対応確定ブロックが得られ、行 L 2 5 と L 2 6 の組み合わせでは、非対応確定ブロックが得られる。

【0098】

類似度がしきい値 T H 1 以上の行については、ステップ S 3 2 は Y E S 側に分岐して、その行の組み合わせに含まれる各ブロックを、対応確定ブロックに決定し、前記構造情報表の対応ブロック番号欄に該当するブロック番号（対応ブロック番号）を記述する（S 3 3）。前記しきい値 T H 1 が 40% であれば、例えば、前記行 L 2 1 では、改版原文文章 O R 2 上のブロック番号 10 のブロックと、旧版原文文章 O R 1 上のブロック番号 8 のブロックが、対応確定ブロックとされる。そして、図 4（A）の構造情報表において、下から 4 番目の行であるブロック番号 8 の行の対応ブロック番号及び類似度の欄には、ブロック番号 10 と、類似度 100% が記述され、同様に、図 4（B）の構造情報表において、もっとも下の行であるブロック番号 10 の行の対応ブロック番号及び類似度の欄には、ブロック番号 8 と、類似度 100% が記述される。

【0099】

非対応確定ブロックに関しては、その対応ブロック番号及び類似度の欄に何も記述しなくてもよいが、必要に応じて、非対応確定ブロックであることを示す所定の情報（非対応確定情報）を記述するようにしてもよい。その場合、前記しきい値 $TH1$ が 40% であるなら、図 17 上の行 $L24 \sim L26$ 内の組み合わせのブロック（図示されていない類似度 0 の組み合わせのブロックも含む）の対応ブロック番号及び類似度の欄には、当該非対応確定情報が記述されることになる。

【0100】

なお、例えば、旧版原文文章 $OR1$ 側のあるブロックに対し、類似度がしきい値 $TH1$ 以上となる改版原文文章 $OR2$ 側のブロックが複数存在すること等も起こり得るが、そのようなケースでは、類似度が最大となるブロックを選択し、選択したブロックを対応確定ブロックとすればよい。

【0101】

当該ステップ $S33$ で行 $L21$ の類似度がしきい値 $TH1$ 以上であることが判明したときには、当該ステップ $S33$ につづいて、当該行 $L21$ を、図 17 の状態のブロック組み合わせ表から削除し（ $S34$ ）、前記ブロック組み合わせ表に残りの行が存在するか否かを検査する（ $S35$ ）。存在する場合、処理は、前記ステップ $S30$ に戻り、存在しない場合には終了する（ $S36$ ）。

【0102】

前記ステップ $S32$ の検査に、前記係数 ρ を反映させることによって、原文文章 $OR1$ と $OR2$ のあいだの親ブロック間の対応関係で、子ブロックの対応関係が規制され、親ブロックの範囲を越えて子ブロックの対応関係が確定する（子ブロックが対応確定ブロックとされる）可能性を低減することができる。

【0103】

これにより、親ブロック間で対応関係が確定していれば、その親ブロックの子ブロック間でも（その親ブロックと対応していないことが確定している親ブロックの子ブロックに比べ）対応関係が確定しやすくなり、子ブロック内の一部に対応しない文が含まれているケースなどでも、子ブロック間の対応関係が確定しやすい。

【0 1 0 4】

以上の処理により、改版原文文章 O R 2 上のすべてのブロックにつき、対応確定ブロックであるか、非対応確定ブロックであるかが決定されるので、その決定に応じて、前記翻訳処理部 8 や差分情報生成部 4 が動作することが可能になる。

【0 1 0 5】

前記翻訳処理部 8 は、改版原文文章 O R 2 中の対応確定ブロックに対しては、対応する旧版訳文文章 C P 1 中のブロックで置き換えることによってブロック単位（例えば、節単位）の対訳による翻訳を実行することができる。また、改版原文文章 O R 2 中の非対応確定ブロックに対しては、通常の機械翻訳を行ったり、前記非特許文献 1 のように、文の類似度に基づいて、文単位で対訳による翻訳を行うことができる。

【0 1 0 6】

以上のような処理を行うことによって、ブロック単位の置き換えを用いる対訳による翻訳を多用した翻訳処理を実行し、前記改版原文文章 O R 2 に対応する改版訳文文章 C P 2 を得ることが可能である。

【0 1 0 7】

改版訳文文章 C P 2 が得られたあと、あるいは、改版訳文文章 C P 2 を得る過程で、前記出力部 7 のディスプレイ装置に例えば図 7 に示すような画面 M G 1 を表示して、ユーザ U 1 が後編集を行ったり、対訳による翻訳を個別的に指示したりするためのユーザインタフェースを提供することができる。

【0 1 0 8】

画面 M G 1 上には、旧版、改版（新版）、原文、訳文の各ブロックに属する 1 または複数の文の文字列を表示するためのフィールド F 1 1 ～ F 1 4 と、ブロック番号を表示するためのフィールド F 2 1, F 2 2 と、フィールド F 1 1 ～ F 1 4 の表示内容をスクロールさせるためのスクロールバー S C 1, S C 2 と、対応関係決定の根拠となった前記類似度を表示するためのフィールド F 2 3 と、対話部品としての各種のボタン B T 1 ～ B T 5 を備えている。

【0 1 0 9】

ユーザ U 1 がポインティングデバイスなどを操作して「次」ボタン B T 1 を押

し下げると、その時点で、フィールド F 1 2 内に表示されている改版原文文章 O R 2 中のブロックが、次のブロック（ブロック番号が 1 つ大きいブロック）に切り替えられる。反対に、「前」ボタン B T 2 を押し下げると、その時点で、フィールド F 1 2 内に表示されている改版原文文章 O R 2 中のブロックが、前のブロック（ブロック番号が 1 つ小さいブロック）に切り替えられる。

【0 1 1 0】

また、旧版と、改版のあいだで、文の文字列が完全に一致しているものには、直観的なマークが付与される。このマークは、前記補助情報に基づいて表示するものであってよい。ユーザ U 1 は、当該マークをもとに文が完全に一致していることを認識することができる。また、一般的に、マークされている文の割合が多いほどそのまま再利用できる確率が高く、対訳による翻訳の結果に対して後編集の必要性が低いことを意味するため、ユーザ U 1 はマークされている文の割合をもとに、そのブロックに対する後編集の要否などを判断することもできる。

【0 1 1 1】

「コピー」ボタン B T 3 は、ユーザ U 1 が、フィールド F 1 1 と F 1 2 に表示されている旧版原文文章 O R 1 中のブロックと、改版原文文章 O R 2 中のブロックを閲覧することによって、これらのブロックの対応付けが良いと判断した場合に押下するボタンである。この押下により、その時点でフィールド F 1 3 に表示されている旧版訳文文章 C P 1 中のブロックが、改版訳文文章 C P 2 のブロックを表示するためのフィールド F 1 4 にコピーされる。したがって、この「コピー」ボタン B T 3 は、対訳による翻訳をユーザ U 1 が個別的に指示するための部品である。

【0 1 1 2】

すでに改版訳文文章 C P 2 が完成している場合には、フィールド F 1 4 には最初から、改版訳文文章 C P 2 中のブロック（翻訳結果の一部）が表示されることになるが、必要に応じて、当該フィールド F 1 4 内には、訳文を 1 文ずつ表示できるようにしてもよい。

【0 1 1 3】

いずれにしても、ユーザ U 1 による編集作業（後編集）は、主として、フィー

ルド F 1 4 に表示されている翻訳結果に対して実行される。

【 0 1 1 4 】

すでに説明したように、旧版原文文章 O R 1 と旧版訳文文章 C P 1 は文のレベルまで厳密に対応している。同様に、改版原文文章 O R 2 と改版訳文文章 C P 2 も、厳密に対応している。また、厳密さではやや低下するものの、旧版原文文章 O R 1 と改版原文文章 O R 2 も、おおむね対応している。したがって、前記ボタン B T 1 や B T 2 を押下して前記フィールド F 1 2 内に表示されている改版原文文章 O R 2 中のブロックを切り替えれば、基本的に、その切り替えに追従して他のフィールド F 1 2 ～ F 1 4 に表示されるブロックも対応したものに切り替えられる。

【 0 1 1 5 】

このような画面 M G 1 を閲覧するユーザ U 1 は、改版原文文章 O R 1 上のブロックを基準に各文章上の所望のブロックを選択して、後編集の作業を進めることができる。選択により、フィールド F 1 4 上に表示されたブロック（改版訳文文章 C P 2 上のブロック）には、改版によって内容が替わったためにそのままでは適切でない文や単語なども含まれている可能性があるので、後編集では、このような文や単語を探し出し、適切なものに変更することになる。

【 0 1 1 6 】

フィールド F 2 3 に表示される類似度は、後編集の必要性が高いブロックをユーザ U 1 に知らせるための情報となる。例えば、類似度が 1 0 0 % のブロックに対しては、通常、後編集を行う必要はないが、類似度が低ければ（例えば、5 0 % 程度なら）、そのブロックに対し重点的に後編集を行う必要があることがわかる。このような類似度に加え、または、類似度に替えて、前記マークなども含む補助情報を活用すれば、フィールド F 1 4 内の画面の色や反転表示などの直感的な方法で、後編集の必要性をユーザ U 1 に伝えることもできる。

【 0 1 1 7 】

後編集を終え、改版訳文文章 C P 2 中の該当ブロックの内容が確定したら、ユーザ U 1 は、「確定」ボタン B T 4 を押下する。これにより、当該ブロックの内容が確定し、保存される。

【0 1 1 8】

後編集や対訳による翻訳の個別的な指示を終えるときには、ユーザU 1 は、「終了」ボタンB T 5を押下する。これにより、前記旧版文書D C 1内のブロックと同様に、当該改版文書D C 2内の該当ブロックが、前記旧版データベース5に格納される。

【0 1 1 9】

このあと、文書D C 2をさらに改版した新たな改版文書D C 3を翻訳する場合、新たな改版文書D C 3からみると文書D C 2は旧版文書であるから、当該旧版データベース5に格納された文書D C 2の対訳は、その新たな改版文書D C 3に対して対訳による翻訳を行う場合に活用することができる。

【0 1 2 0】

(A-3) 第1の実施形態の効果

本実施形態によれば、文脈に忠実で品質の高い翻訳結果を得ることが可能である。

【0 1 2 1】

また、本実施形態では、文脈に忠実な翻訳を行う過程で得られた各種の情報（前記補助情報なども含む）を活用することにより、後編集の作業効率を高めることができる。

【0 1 2 2】

(B) 第2の実施形態

以下では、本実施形態が第1の実施形態と相違する点についてのみ説明する。

【0 1 2 3】

本実施形態では、文に対して類似度を算出して文間の対応関係を決定する際、隣接する文が対応確定文（対応関係の確定した文）である場合や、近傍の文に対応確定文が多い場合など、その文に対して距離が近い文が対応確定文となったときには、その文の類似度を高める方向に制御することを特徴とする。

【0 1 2 4】

(B-1) 第2の実施形態の構成および動作

構成上、本実施形態が第1の実施形態と相違するのは、図8に示すように、詳

細照合部 3 B に対し類似度重み付け部 3 C が接続されている点だけである。

【0125】

本実施形態の翻訳支援システム 10 における文間の対応関係を決定する際の動作を図 9 のフローチャートに示す。図 9 のフローチャートは、S 40～S 47 の各ステップを備えている。

【0126】

ここで、前記旧版文書 DC 1 に相当する旧版文書を DC 11 とし、前記改版文書 DC 2 に相当する改版文書を DC 21 とする。また、DC 11 内の旧版原文文章 OR 11 の 1 ブロックである BR 1 には、文 a、文 b、文 c、文 d が含まれ、DC 21 内の改版原文文章 OR 21 の 1 ブロックである BR 2 には、文 1 C、文 2 C、文 3 C、文 4 C が含まれているものとする。各文章 OR 11、OR 21 上に各文が出現する順番もここに記載した通りであるものとする。そして、改版原文文章 DC 21 中の文 1 C は、旧版原文文章 DC 11 中の文 a を 1 文字も変更することなくそのまま用いており、残りの文 2 C～4 C は、改版により、変更または追加されたものと仮定する。

【0127】

また、前記ステップ S 40 が実行されるまえに、各文章 OR 11 と OR 21 上におけるブロック間の対応関係はすでに決定されているものとし、図 9 では、各ブロック内における文の対応関係を決定する。

【0128】

図 9 において、改版原文文章 OR 21 と旧版原文文章 OR 11 のあいだで対応関係が確定している対応確定ブロックを 1 つずつ選択する (S 40)。これにより、例えば、前記ブロック BR 1 と BR 2 が選択される。

【0129】

次に、当該ブロック BR 1、BR 2 のあいだで、すべての文字が一致している文の組み合わせを選定し (S 41)、選定した組み合わせに含まれる文以外の文に対して単語の切り出し処理を行う (S 42)。このステップ S 41 では、前記文 1 C と文 a の組み合わせが選定される。文 1 C と文 a の組み合わせについては、この時点で、対応関係があることが確定し、改版原文文章 OR 21 上で当該文 1

Cが、前記対応確定文とされる。

【0130】

また、ステップS42の単語の切り出し処理は、例えば、形態素解析によって行うことができるが、必要ならば、単語の切り出しの代わりに文字の切り出しを行うようにしてもよい。

【0131】

ここで、単語の切り出しを行うのは、後述する式(2)で類似度を算出するためである。

【0132】

ステップS42につづくステップS43では、ブロックBR2のなかで対応関係が決定されていない文を1文ずつ選び、ブロックBR1中の各文に対するすべての組み合わせ(すべての文の対)で、次の式(2)に基づく重み付け類似度(修正類似度)を算出する。

【0133】

$$WT \times 100 \times \text{一致単語数} / ((1 \text{ 対の文の総単語数}) / 2) \quad \dots (2)$$

ここで、WTは重みで、その初期値は1である。ただし、該当する文章(ここでは、OR21)上でその文の1文前または1文後に出現する文の対応関係が決定されている場合には、重みWTの値は初期値より大きくなるように変更される。初期値の次の値は、例えば、1.2であってよい。同様な重みWTの値の変更は繰り返され、近傍に出現する対応確定文の密度が高いほど、WTの値が大きくなるように変更したり、これと反対に、近傍に、対応する文が存在しない旨の決定が行われた文(非対応確定文)が出現し、その密度が高くなるほど、WTの値が小さくなるように変更するものであってもよいが、図10(A)～(C)の例では、重みWTの値は、初期値の1と1.2の2通りであるものとする。また、密度などは考慮せず、単純に隣接する文の対応関係が確定したときには重みWTの値を1から1.2に変化させるものとする。

【0134】

同様な重み付け類似度は、すでに対応関係が決定されたもの(例えば、前記文aと文1Cの組み合わせ等)を除き、前記ブロックBR1, BR2間で可能なす

すべての組み合わせについて算出される。

【0 1 3 5】

もしも、前記文 2 C と文 b の具体的な文字列が次の通りであり、重み W T の値が 1 であるものとする、文 2 C の単語数は 5 で、文 b の単語数は 6 であり、文 2 C と文 b から構成される対の文の総単語数は 1 1 である。

【0 1 3 6】

文 2 C : This is a pencil .

文 b : This is a pencil case .

この場合、一致単語数は、5 であるから、前記式 (2) によって得られる重み付け類似度は、9 0 . 9 % ($\div 1 \times 1 0 0 \times 5 / (1 1 / 2)$) となる。

【0 1 3 7】

次に、重み付け類似度が所定のしきい値 T H 1 以上となる組み合わせを選定する (S 4 4) 。しきい値 T H 1 の具体値は、第 1 の実施形態と同じであってもよいが、異なるものであってもよい。ここでは、例えば、5 0 % であるものとする。なお、旧版原文文章 O R 1 1 側のある文に対する改版原文文章 O R 2 1 側の複数の文の組み合わせの重み付け類似度が、同時に、前記しきい値 T H 1 以上となることは起こり得るが、そのようなケースでは、重み付け類似度が最大の組み合わせだけについて、対応関係を決定するとよい。

【0 1 3 8】

前記文 2 C ~ 4 C と文 b ~ d の各組み合わせについて算出した重み付け類似度が、例えば、図 1 0 (A) に示す通りであるものとする、文 b と文 2 C の組み合わせの重み付け類似度 (ここでは、5 6 . 4 %) だけがしきい値 T H 1 以上であるため、その組み合わせの対応関係が決定され、文 2 C が対応確定文とされる。

【0 1 3 9】

そのブロック B R 2 内に対応関係が確定していない文があり、なおかつ、その回のループ (ステップ S 4 3 ~ S 4 6 によって構成されるループ) の処理によって、新たに、対応確定文とされた文がある限り、ステップ S 4 3 ~ 4 6 の処理は繰り返される。

【0 1 4 0】

この繰り返しのたびに、異なる文が対応確定文となるため、値 1. 2 の重み W_T が反映される文が変化する。例えば、図 1 0 (A) ~ (C) の例において、図 1 0 (A) では、すでに対応確定文となった文 1 C に隣接している文 2 C に、値 1. 2 の重み W_T が使用され、重み W_T の値が 1 ならば 4 7 となる類似度が、重み W_T の値が 1. 2 となったことによって、5 6. 4 (重み W_T が 1 ならば、4 5) になり、しきい値 TH_1 (= 5 0) 以上に達している。

【0 1 4 1】

同様に、図 1 0 (B) でも、当該文 2 C が対応確定文となったことにより、文 2 C に隣接する文 3 C が値 1. 2 の重み W_T の影響を受け、重み付け類似度が 5 4 になり、しきい値 TH_1 以上に達し、対応確定文とされる。

【0 1 4 2】

最後の図 1 0 (C) でも、当該文 3 C が対応確定文となったことにより、文 3 C に隣接する文 4 C が値 1. 2 の重み W_T の影響を受け、重み付け類似度が 4 8 になるが、4 8 では、前記しきい値 TH_1 以上ではないため、当該文 4 C と文 d の組み合わせは、対応関係がないことが決定され、文 4 C は非対応確定文とされる。

【0 1 4 3】

以上と同様な処理は、改版原文文章 OR 2 1 上のすべてのブロックについて実行される (S 4 7)。

【0 1 4 4】

(B) 第 2 の実施形態の効果

本実施形態によれば、第 1 の実施形態の効果と同等な効果を得ることができる。

【0 1 4 5】

加えて、本実施形態では、対応確定文に対する距離が近い (隣接する) 文は、その重みの値が大きくなるため対応確定文とされやすくなる。これにより、1 文だけで見ると他に類似度の高い文があったとしても、前後の文が編集されていないか編集量が少ないと対応確定文とされやすくなり、連続して対応確定文が発生する傾向が高くなる。これは、文脈に忠実な翻訳結果を得るために有効である。

【0 1 4 6】

反対に、改版により、ある文に隣接する文が削除されるか大幅に編集された場合には、隣接する文の類似度は相対的に小さくなるため、その文は、対応確定文となりにくい。このようなケースでは、その削除や編集により、文のつながりが弱まっているといえるから、この意味でも、本実施形態は、文脈に忠実な翻訳結果を得やすいといえる。

【0 1 4 7】

(C) 第3の実施形態

以下では、本実施形態が第1、第2の実施形態と相違する点についてのみ説明する。

【0 1 4 8】

本実施形態は、ユーザインタフェースが第1の実施形態と相違し、いっそう後編集を行いやすくなっている。

【0 1 4 9】

(C-1) 第3の実施形態の構成および動作

構成上、本実施形態が第1、第2の実施形態と相違するのは、主として、図11に示すように、前記画面MG1に対応する画面MG2上に「情報」ボタンBT6を設けた点である。「情報」ボタンBT6は、ユーザU1が後編集のための情報の提供を要求する際に押下する。

【0 1 5 0】

本実施形態の翻訳支援システム10における画面表示のための動作を図12のフローチャートに示す。図12のフローチャートは、S50～S53の各ステップを備えている。

【0 1 5 1】

図12において、図11の画面MG2上の改版文章上のブロックが表示されるフィールドF12またはF14（必要に応じて、F11、F13としてもよい）上に所望のブロック（子ブロック）が表示された状態で、ユーザU1が「情報」ボタンBT6を押下すると、その時点でフィールドF21に表示されていたブロック番号が制御部6に供給され、制御部6は、当該ブロック番号で指定されるブ

ロックの上位ブロック（親ブロック）のブロック番号を探索する（S 5 0）。この探索は、例えば、図 4（A）および（B）に示す構造情報表を利用することで、容易に実行可能である。

【0 1 5 2】

当該親ブロックは、対応確定ブロックである場合と、非対応確定ブロックである場合があり得るが、非対応確定ブロックである場合にはステップ S 5 1 が N O 側に分岐して、当該親ブロックが非対応確定ブロックである旨が前記ディスプレイ装置の画面（図示せず）を介してユーザ U 1 に伝えられるだけである。これは、当該親ブロックが改版によって新たに追加されたブロックであるケースなどに発生する。

【0 1 5 3】

一方、当該親ブロックが対応確定ブロックである場合には、ステップ S 5 1 は Y E S 側に分岐し、同じ親ブロックを親とする改版文章側の他の子ブロック（並列ブロック）を探索する（S 5 2）。ここで、改版文章は、改版原文文章としてもよいが、後編集の性質上、改版訳文文章とするのが普通であると考えられる。同様な探索は、その親ブロックと対応関係が確定された旧版文章側でも行われ、改版文章と旧版文章のあいだで、各子ブロックのあいだの関係（相互に対応確定ブロックであるか、非対応確定ブロックであるか）が調べられ、相互に対応確定ブロックである場合には、対応確定ブロックの根拠とされた類似度が表示される。そのためにディスプレイ装置上に表示される画面は、例えば、図 1 3 に示す画面 M G 6 のような構成であってよい。

【0 1 5 4】

画面 M G 6 上には、基本的に前記並列ブロックを表示するが、必要に応じて、同じ親ブロックを親としない子ブロックが表示されるものであってもよい。図 1 3 の例では、後述するように、ブロック A 5 がそのような子ブロックである。

【0 1 5 5】

図 1 3 において、A 1 ～ A 5 は、旧版文章側の子ブロックであり、B 1 ～ B 6 は改版文章側の子ブロックである。また、画面 M G 3 上でブロック間を接続する対応線 N K 1 ～ N K 5 は、接続した各ブロックが相互に対応関係が確定した対応

確定ブロックであることを直観的に示している。各対応線NK1～NK5の近傍に表示した数字（100，50，80など）は、対応関係確定の根拠となった類似度である。

【0156】

一般的に類似度の値が小さいほど改版による変更の割合が高く後編集の必要性が高いので、ユーザU1は、表示された当該類似度をもとに後編集するブロックを選ぶことができ、類似度の低いものに注力して効率的な後編集を行うことができる。

【0157】

さらに、画面MG3から対応確定ブロックの旧版、改版文章上における位置関係（並び方）がわかるため、その位置関係に基づいて、さらに詳細に後編集の対象を選ぶことができる。例えば、ブロックB2の場合、1ブロック前のB1とA1は対応していることから、ブロックB2の前半部分は後編集の必要性が低いと判断することができるが、1ブロック後ろのブロックB3とA3は対応していないため、ブロックB2の後半部分は後編集の必要性が高いと判断できる。

【0158】

なお、いずれの対応線によっても接続されないブロックB5は、改版によって新規に追加したと判断されたブロックである。また、図13上で他のブロックよりも太い線で表示されたブロックB2，A2は、前記「情報」ボタンBT6を押下する前に画面MG2のフィールドF14内に表示していた子ブロックである。このような表示により、ユーザU1は、後編集作業中に、最初に注目していた子ブロック（B2）を見失うことがなくなる。

【0159】

さらに、対応線が実線ではなく、点線となっているNK5は、親ブロックは対応関係がないブロックである。すなわち当該ブロックA5は旧版文章上で他のブロックA1～A4とは異なる親ブロックの子ブロックとなっている。このようなケースでは、対訳による翻訳結果としての当該ブロックB6は、文脈に忠実なものとはなっていない可能性が高いため、類似度こそ80%と比較的高いものの、当該ブロックB6は、後編集の必要性が高いと判断できる。

【0 1 6 0】

なお、図 1 3 では、各ブロック内には何も表示されていないが、必要に応じて、具体的な文字列の内容を表示するようにしてもよい。例えば、各ブロック内にそのブロックに属する最初の 1 文を表示することも望ましい。

【0 1 6 1】

再度、前記画面 MG 2 を表示させ、画面 MG 2 上で各フィールド F 1 1 ~ F 1 4 に表示するブロックを変更した上で前記「情報」ボタン BT 6 を押下すれば、異なる階層でも図 1 2 のフローチャートの処理を行うことができることは当然である。

【0 1 6 2】

(C-2) 第 3 の実施形態の効果

本実施形態によれば、第 1、第 2 の実施形態と同等な効果を得ることができる。

【0 1 6 3】

加えて、本実施形態では、該当子ブロック（例えば、B 2）が属する上位ブロック（B 1 ~ B 4 を子ブロックとする親ブロックなど）内の全範囲におよぼ変更情報（例えば、対応線 NK 1 ~ NK 4（NK 5）やその近傍に表示した類似度など）を表示することができるため、旧版文章と改版文章の全体的な差分がわかりやすくなり、文脈にあった後編集作業が行いやすくなる。

【0 1 6 4】

また改版による変更の影響がどのように波及しているかを直観的に概観できるため、後編集に要する時間を見積もることも可能になる。

【0 1 6 5】

(D) 第 4 の実施形態

以下では、本実施形態が第 1 ~ 第 3 の実施形態と相違する点についてのみ説明する。

【0 1 6 6】

第 1 ~ 第 3 の実施形態では、ブロック間の対応関係は翻訳支援システムが自動的に決定するものであったが、本実施形態では、翻訳支援システムが自動的に確

定したブロック間の対応関係（対応確定ブロック）をユーザU 1 が検証し、必要に応じて、ユーザU 1 が対応関係の変更を行うことが可能な構成となっている。

【0 1 6 7】

（D-1）第4の実施形態の構成および動作

構成上、本実施形態が第1～第3の実施形態と相違するのは、主として、図14に示す画面MG4である。画面MG4は、前記画面MG1に対応する画面であるが、「次候補」ボタンBT7と「前候補」ボタンBT8を備えている点が相違する。

【0 1 6 8】

「次候補」ボタンBT7と「前候補」ボタンBT8は、ユーザU 1 が対応確定ブロックを変更する際、新たな対応確定ブロックの選択を行うためのボタンである。旧版文章側のあるブロックに対応する改版文章側のブロックは、その類似度の高さをもとに整列された形で、ブロック対応表として翻訳支援システム10内に蓄積されている。

【0 1 6 9】

ブロック対応表は、例えば、図17に示したブロック組み合わせ表と同様な表であってもよいが、類似度が前記しきい値TH1以上のブロックの組み合わせのみを格納した表である。また、図17のブロック組み合わせ表は、同じ階層位置における任意の組み合わせを、単に類似度の高さに応じて整列しただけの表であるが、ブロック対応表の場合、旧版文章側の各ブロックごとに整理し、改版文章側の各ブロックを、類似度の高さに応じて整列したものであることが望ましい。

【0 1 7 0】

ただし、図17の表も、当該表に対する検索条件の生成のしかた次第で、ブロック対応表として活用することが可能である。

【0 1 7 1】

要するに、旧版文章側のあるブロックに対して対応関係をなす改版文章側のブロックの候補（候補ブロック）を複数用意しておき、ユーザU 1 からの指示に応じて、候補ブロックを選択し、ブロックの組み合わせを変更できる構成であればよい。

【0 1 7 2】

第 1 の実施形態では、図 6 に示したフローチャートのステップ S 3 3 で構造情報表に対応ブロック番号を記述するとき、例えば、旧版原文文章 O R 1 側のあるブロックに対し、類似度がしきい値 T H 1 以上となる改版原文文章 O R 2 側のブロックが複数存在する場合には、類似度が最大となるブロックを選択して対応確定ブロックとしたが、本実施形態では、この選択によって選ばれなかったブロックのブロック番号も、候補ブロック番号として保存しておくことになる。

【0 1 7 3】

図 1 4 に示す画面 M G 4 を閲覧したユーザ U 1 が、前記「次候補」ボタン B T 7 を押下すると、その時点で、例えば、前記フィールド F 2 2 に表示されていたブロック番号が制御部 6 に供給され、当該ブロック番号をもとに制御部 6 が、前記ブロック対応表に対する検索を行い、検索結果として類似度の高さが 2 番目以降のブロックのブロック番号を取得する。そして、そのブロック番号に対応したブロックの本体を前記旧版データベース 5 から取得して、画面 M G 4 の該当するフィールド（例えば、F 1 2）に表示する。このとき、当該ブロックのブロック番号が、前記フィールド（例えば、F 2 2）に表示されることになる。

【0 1 7 4】

以降は同様な処理が繰り返され得る。

【0 1 7 5】

ユーザ U 1 は、「次候補」ボタン B T 7 を押下するたびに類似度がより低い候補ブロックを閲覧することができ、前候補ボタン B T 8 を押下するたびに類似度がより高い候補ブロック（もとの対応確定ブロックも含む）を閲覧することができるため、いずれのブロックが、その対応確定ブロックとして最適であるかを、ユーザ U 1 自身で判断することができる。

【0 1 7 6】

ユーザ U 1 の判断によって、対応確定ブロックが変更された場合には、改版訳文文章 C P 2 の内容も変更されることになる。

【0 1 7 7】

(D-2) 第 4 の実施形態の効果

本実施形態によれば、第1～第4の実施形態と同等な効果を得ることができる。

【0178】

加えて、本実施形態では、翻訳支援システム（10）が自動的に確定したブロック間の対応関係をユーザ（U1）が検証し、必要に応じて、ユーザ（U1）が対応関係の変更を行うことも可能である。これは、翻訳支援システム（10）の使い勝手を高めるとともに、対訳による翻訳結果の品質向上にも寄与する。

【0179】

（E）他の実施形態

上記第1～第4の実施形態では、多くの画面の具体的な構成例を図示したが、図示した以外の構成を持つ画面を用いてもよいことは当然である。

【0180】

なお、上記第2の実施形態では、主として、隣接する文が対応確定文となったときにその文の類似度を高める場合を例に説明したが、この処理を、近傍の文に対応確定文が多い場合や、その文に対して距離が近い文が対応確定文となった場合にその文の類似度を高めるように拡張することは容易である。

【0181】

また、上記第1～第4の実施形態では、段落のブロックを無視したが、段落に配慮した処理を行ってもよいことは当然である。

【0182】

なお、上記第2の実施形態で述べた文は、ブロックに置換可能である。すなわち、隣接するブロックが対応確定ブロックである場合や、近傍のブロックに対応確定ブロックが多い場合に、ブロックの類似度を高める方向に制御するようにしてもよい。

【0183】

また、上記第1～第4の実施形態にかかわらず、翻訳を行うことは必ずしも必須ではない。本発明は、ブロックの対応関係を検出して、文脈を利用し、マニュアルなどの詳細な版管理を行う場合など（旧版文書と改版文書の詳細な差分に関する情報の取得なども含む）にも適用することが可能である。また、版管理以外

で、文書間でブロックの対応関係を検出する場合にも適用することが可能である。

【0184】

さらに、文書は、自然言語以外の構成要素を含んでいてもかまわない。例えば、図形や、画像などを含む文書に対しても本発明は適用可能である。また、図形や画像なども文書内における文脈の形成に寄与し得ることは当然である。

【0185】

また、文書は自然言語以外の言語（例えば、プログラミング言語など）を含むものであってもかまわない。プログラミング言語で記述されるコンピュータプログラムのソースコードの文書は、前記マニュアル、技術文書、論文などと同様、しばしば改版を行う文書の代表例の一つである。

【0186】

以上の説明では主としてハードウェア的に本発明を実現したが、本発明はソフトウェア的に実現することも可能である。

【0187】

【発明の効果】

以上に説明したように、本発明によれば、文書と文書の対応関係を、その文脈に配慮して検出することができる。

【0188】

これにより、例えば、版管理の品質を高めることや、対訳辞書を用いた翻訳処理の品質を高めることも可能になる。

【図面の簡単な説明】

【図1】

第1の実施形態にかかる翻訳支援システムの全体構成例を示す概略図である。

【図2】

第1～第4の実施形態で処理する原文文章の構成例を示す概略図である。

【図3】

第1の実施形態の動作例を示すフローチャートである。

【図4】

第1～第4の実施形態で使用する原文文章の階層構造の例を示す表である。

【図5】

第1の実施形態の動作例を示すフローチャートである。

【図6】

第1の実施形態の動作例を示すフローチャートである。

【図7】

第1の実施形態の動作説明図である。

【図8】

第2の実施形態にかかる翻訳支援システムで使用する文書構造比較部の

【図9】

第2の実施形態の動作例を示すフローチャートである。

【図10】

第2の実施形態の動作説明図である。

【図11】

第3の実施形態の動作説明図である。

【図12】

第3の実施形態の動作例を示すフローチャートである。

【図13】

第3の実施形態の動作説明図である。

【図14】

第4の実施形態の動作説明図である。

【図15】

第1～第4の実施形態の動作説明図である。

【図16】

第1～第4の実施形態の動作説明図である。

【図17】

第1～第4の実施形態の動作説明図である。

【図18】

第1～第4の実施形態の動作説明図である。

【図 1 9】

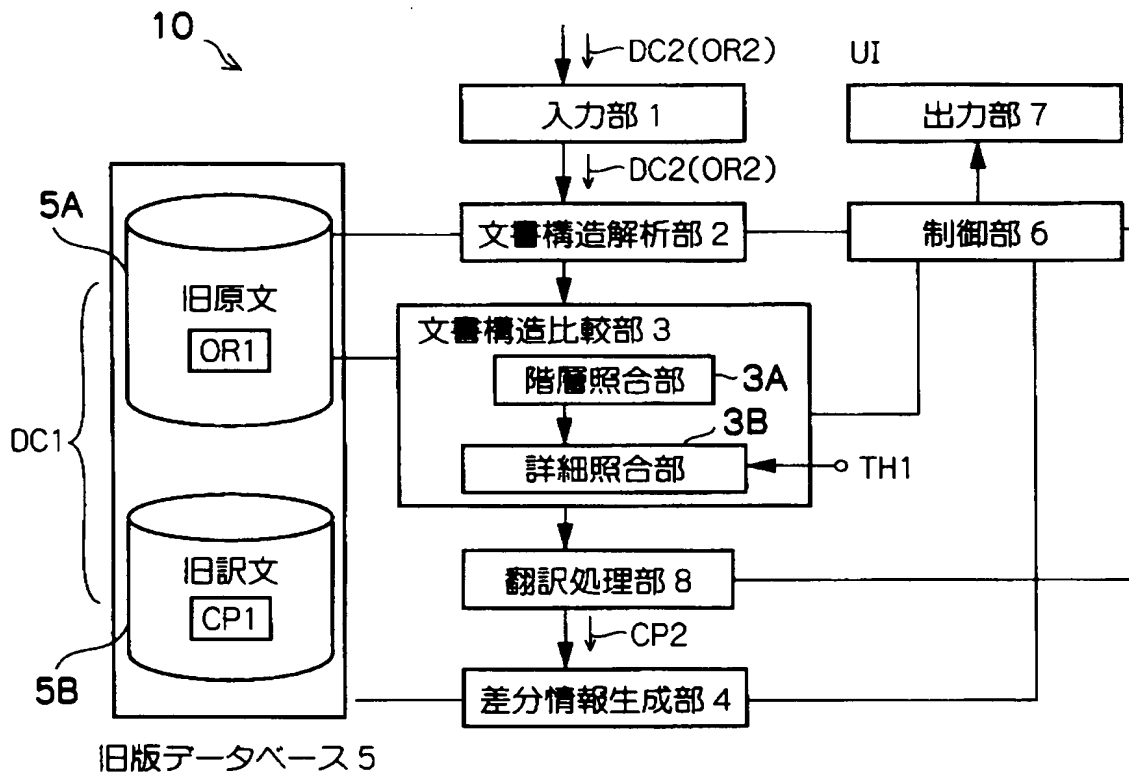
第 1 ～ 第 4 の実施形態の動作説明図である。

【符号の説明】

1 …入力部、2 …文書構造解析部、3 …文書構造比較部、3 A …階層照合部、3 B …詳細照合部、4 …差分情報生成部、5 …旧版データベース、5 A …旧版原文データベース、5 B …旧版訳文データベース、6 …制御部、7 …出力部、8 …翻訳処理部、O R 1 …旧版原文文章、O R 2 …改版原文文章、C P 1 …旧版訳文文章、C P 2 …改版訳文文章、D C 1 …旧版文書、D C 2 …改版文書、M G 1 ～ M G 4 …画面。

【書類名】 図面

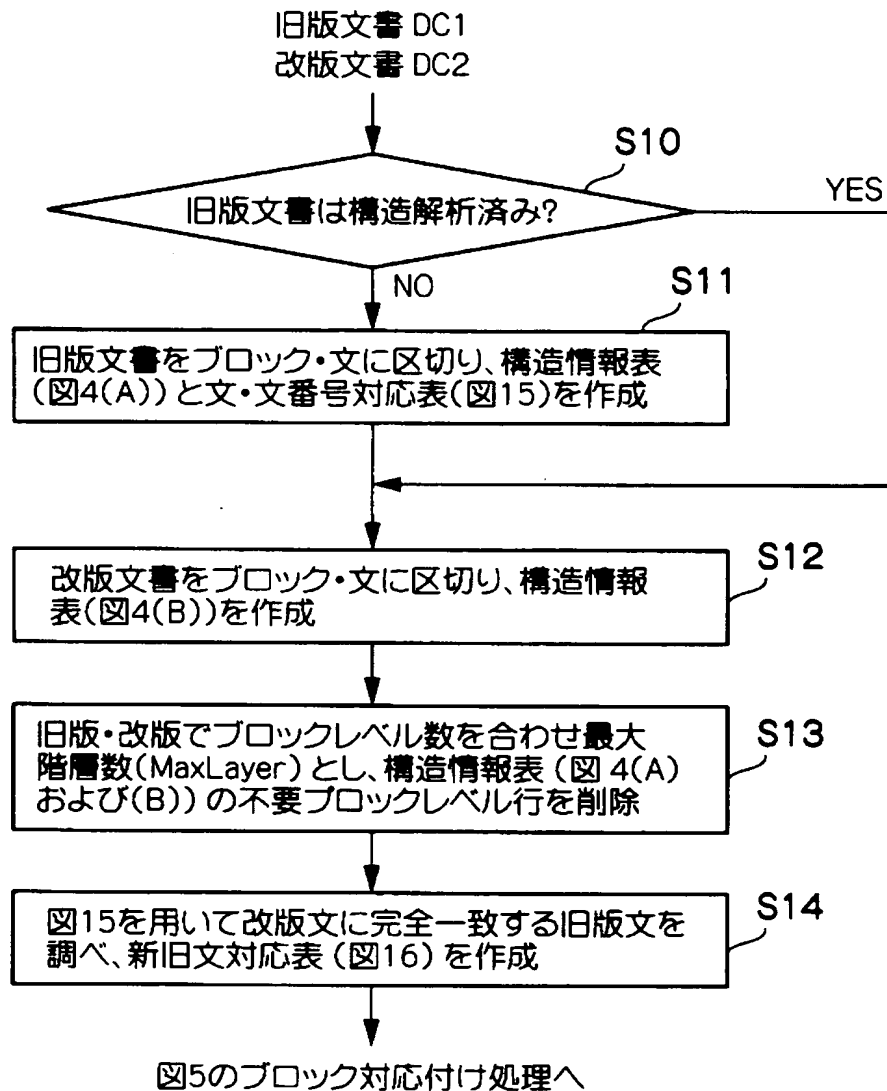
【図 1】



【図 2】

(A) (a)旧版原文文章 OR1	(B) (b)改版原文文章 OR2
<u>1</u> <u>1.1</u> 文1。文2。文3。文4。 <u>1.2</u> 文5。文6。文7。 <u>2</u> <u>2.1</u> 文8。文9。文10。文11。 <u>2.2</u> 文12。文13。文14。 <u>2.3</u> 文15。 <u>3</u> <u>3.1</u> 文16 <u>3.2</u> 文17。 <u>4</u> 文6。	<u>1</u> <u>1.1</u> 文A (1)。文B (2)。文C (3)。 文D(4)。 <u>1.2</u> 文E(6)。文F(7)。文G14)。 <u>2</u> 文H(15)。 <u>3</u> <u>3.1</u> 文I(12)。文J(13)。 <u>3.2</u> <u>3.2.1</u> 文K(9)。文L(10)。 <u>3.2.2</u> 文M (18)。文N(19)。文O(20)。 <u>4</u> 文P (16)、文Q(17)

【図 3】



【図 4】

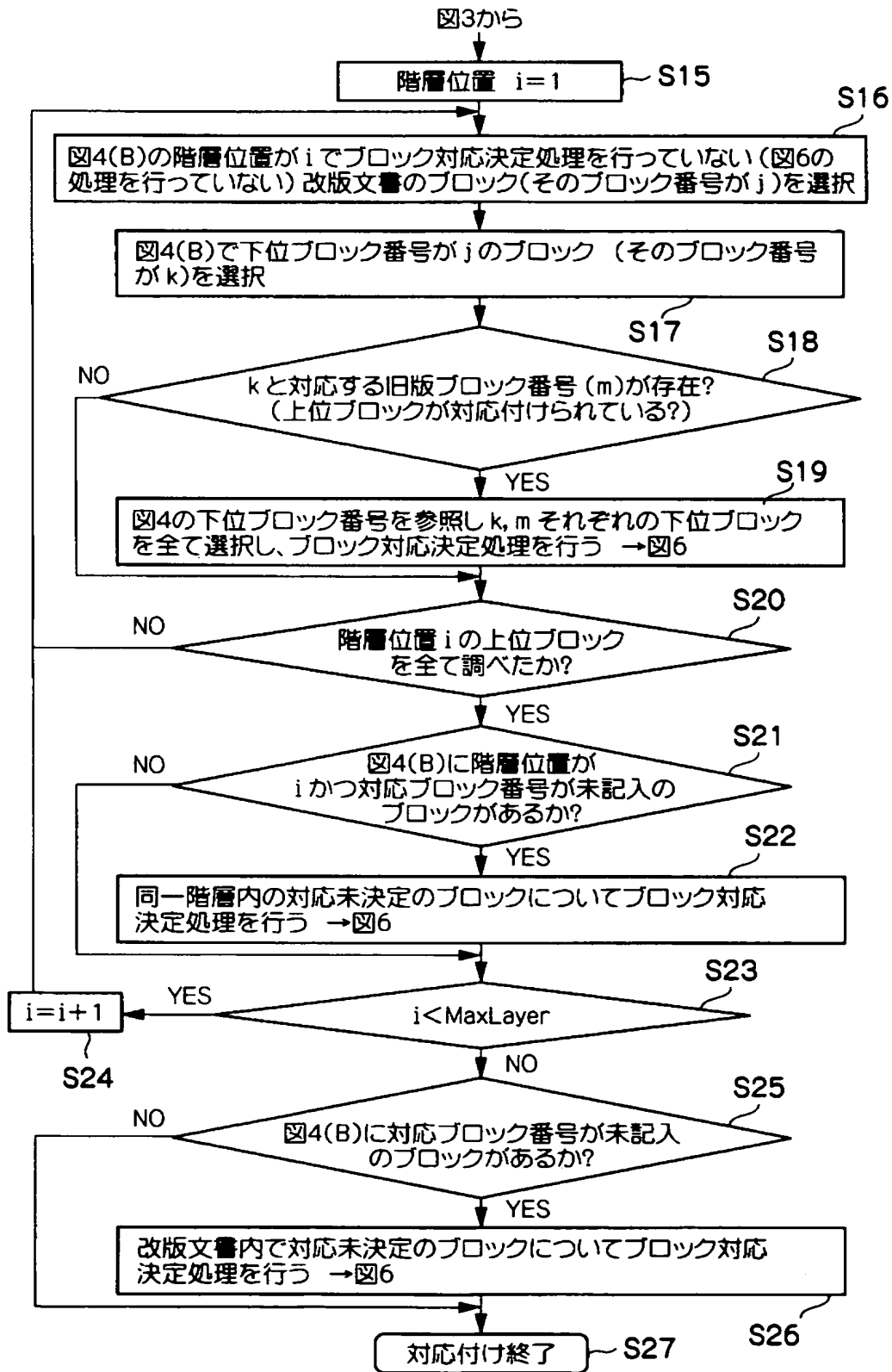
(A)旧版原文文章 OR1 の階層構造

ブロック 番号	階層 位置	下位ブロック番号	文 番 号	対応ブロック 番号及び 類似度
0	0	1、4、8、11	1~17	0
1	1	2、3	1、2、3、4、5、6、7	
2	2		1、2、3、4	
3	2		5、6、7	
4	1	5、6、7	8、9、10、11、12、13 14、15	
5	2		8、9、10、11	
6	2		12、13、14	
7	2		15	
8	1	9、10	16、17	
9	2		16	
10	2		17	
11	1		6	

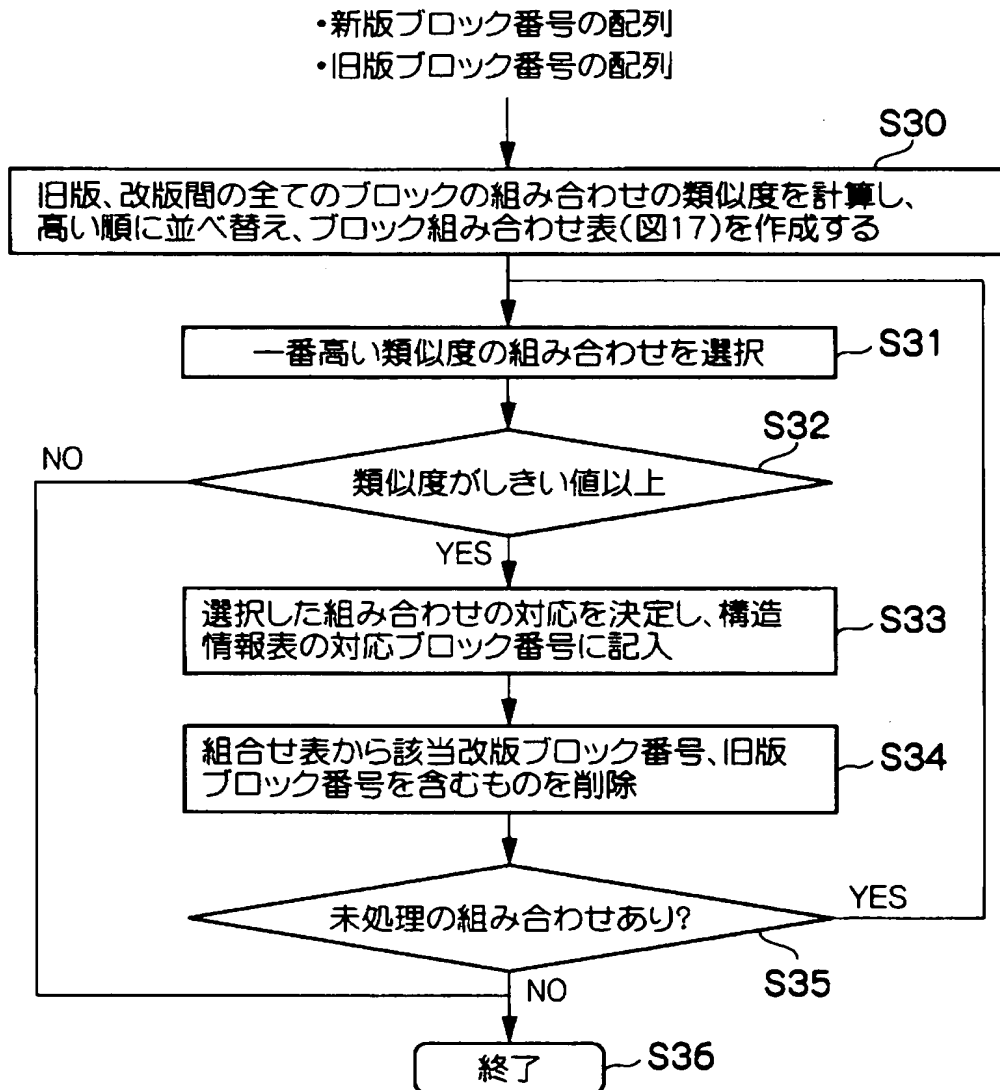
(B)改版原文文章 OR2 の階層構造

ブロック 番号	階層 位置	下位ブロック番号	文 番 号	対応ブロック 番号及び 類似度
0	0	1、4、5、10	A ~ Q	0
1	1	2、3	A,B,C,D,E,F,G	
2	2		A,B,C,D	
3	2		E,F,G	
4	1		H	
5	1	6、7	I,J,K,L,M,N,O	
6	2		I,J	
7	2	8、9	K,L,M,N,O	
8	3		K,L	
9	3		M,N,O	
10	1		P,Q	

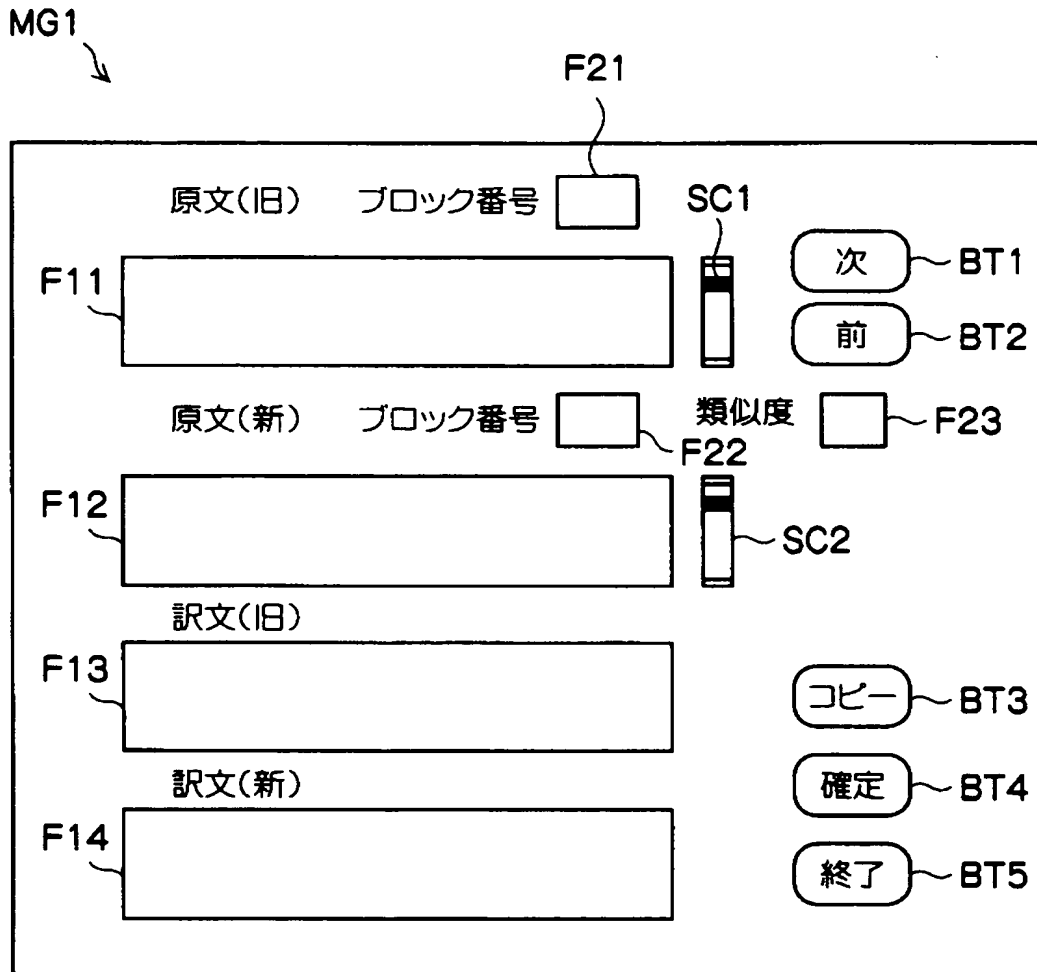
【図 5】



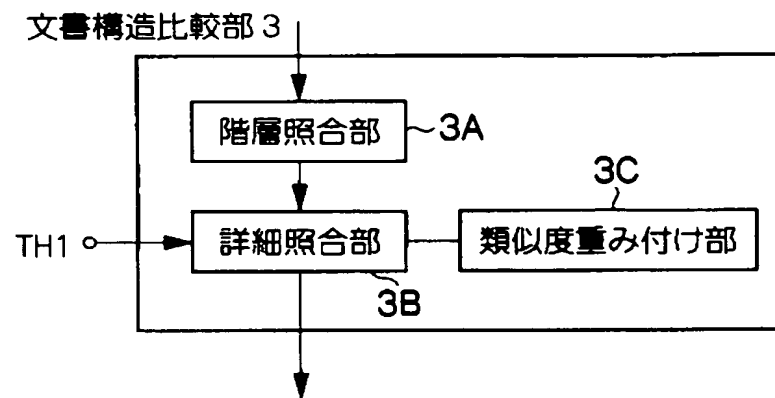
【図 6】



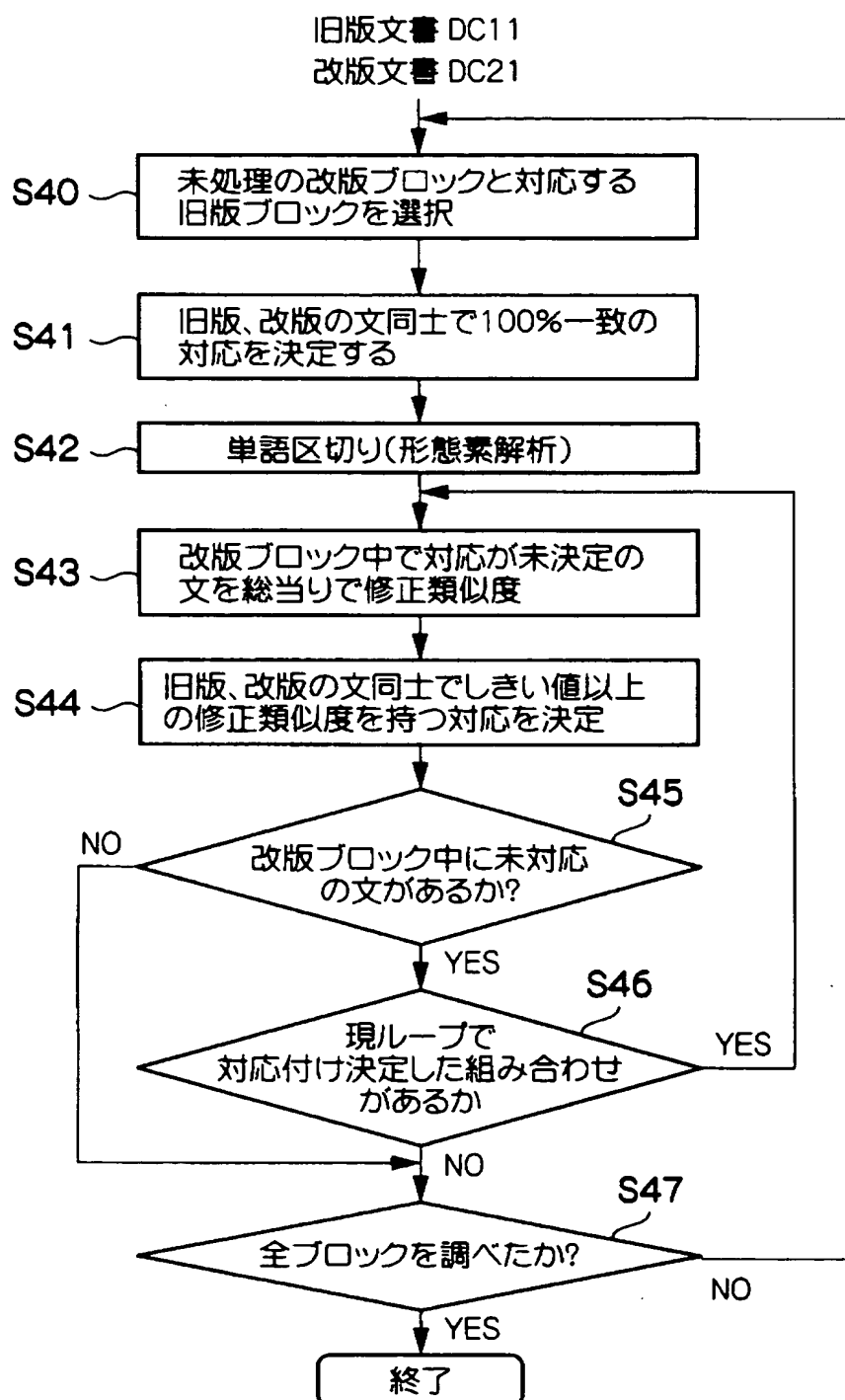
【図 7】



【図 8】



【図 9】



【図 10】

原文の重み付け類似度(1回目)

(A)		文 2C	文 3C	文 4C
	文 b	56.4(47)	40	50
	文 c	16	45	22
	文 d	20	35	40

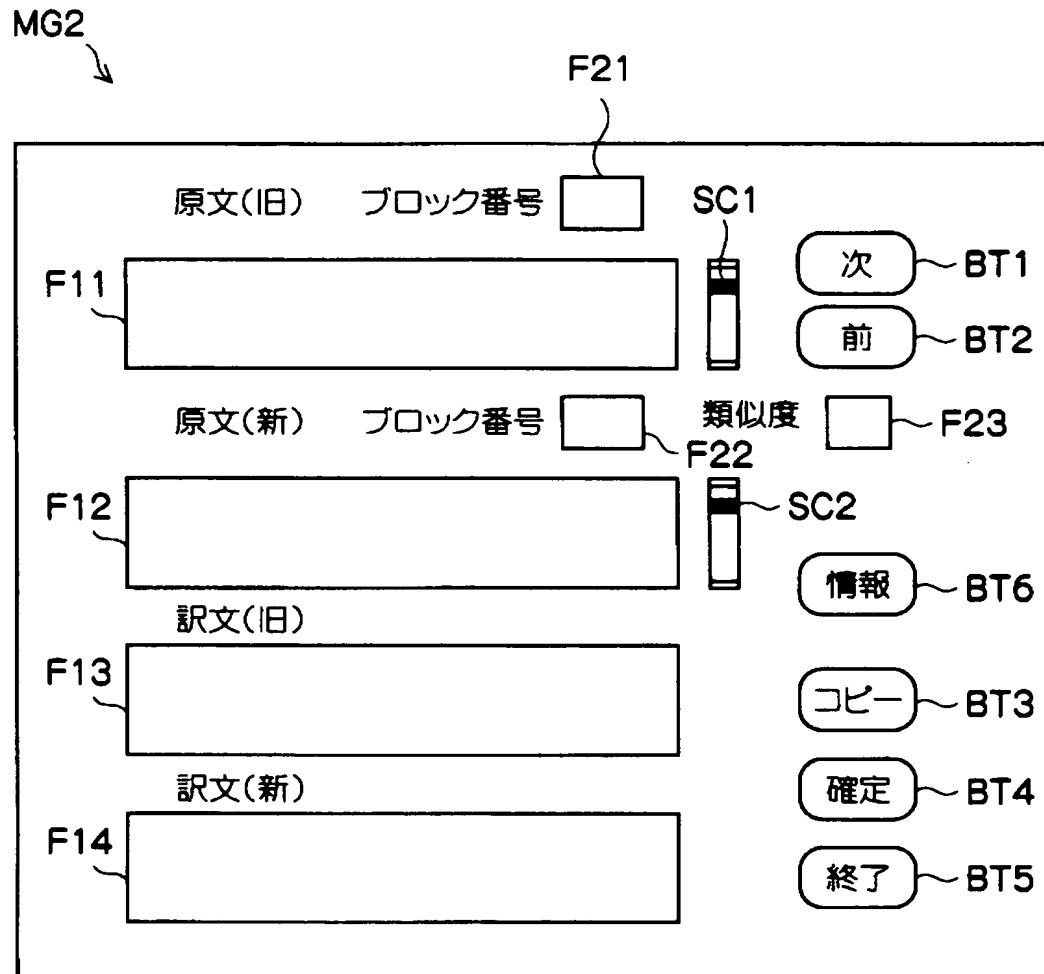
重み付け類似度(2回目)

(B)		文 3C	文 4C
	文 c	54(45)	22
	文 d	35	40

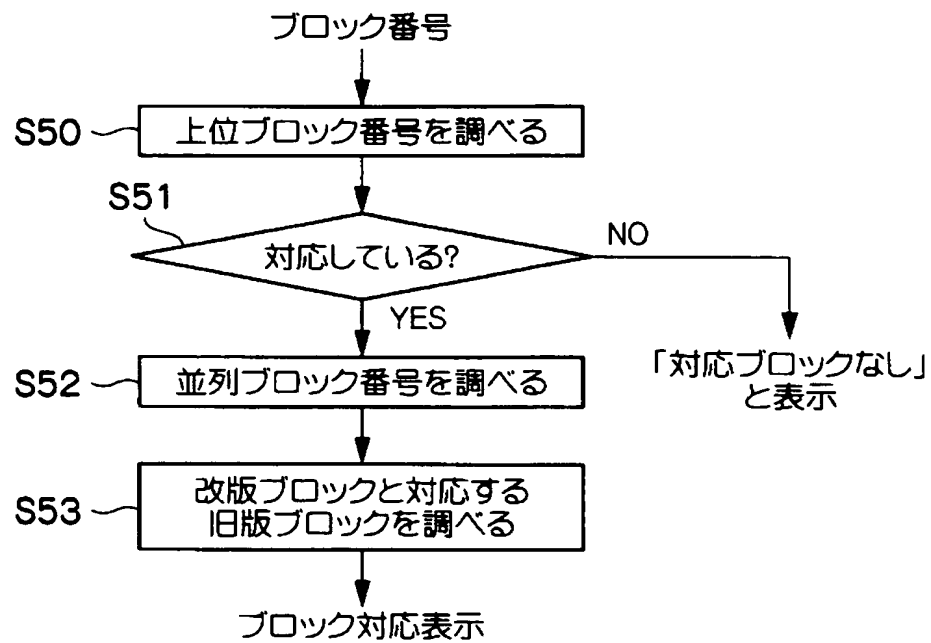
重み付け類似度(3回目)

(C)		文 4C
	文 d	48(40)

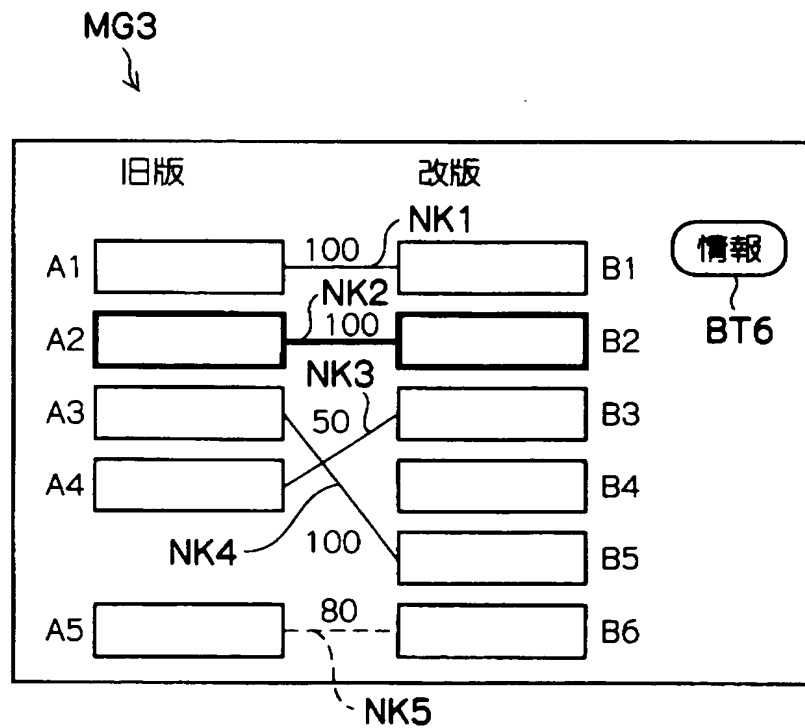
【図 11】



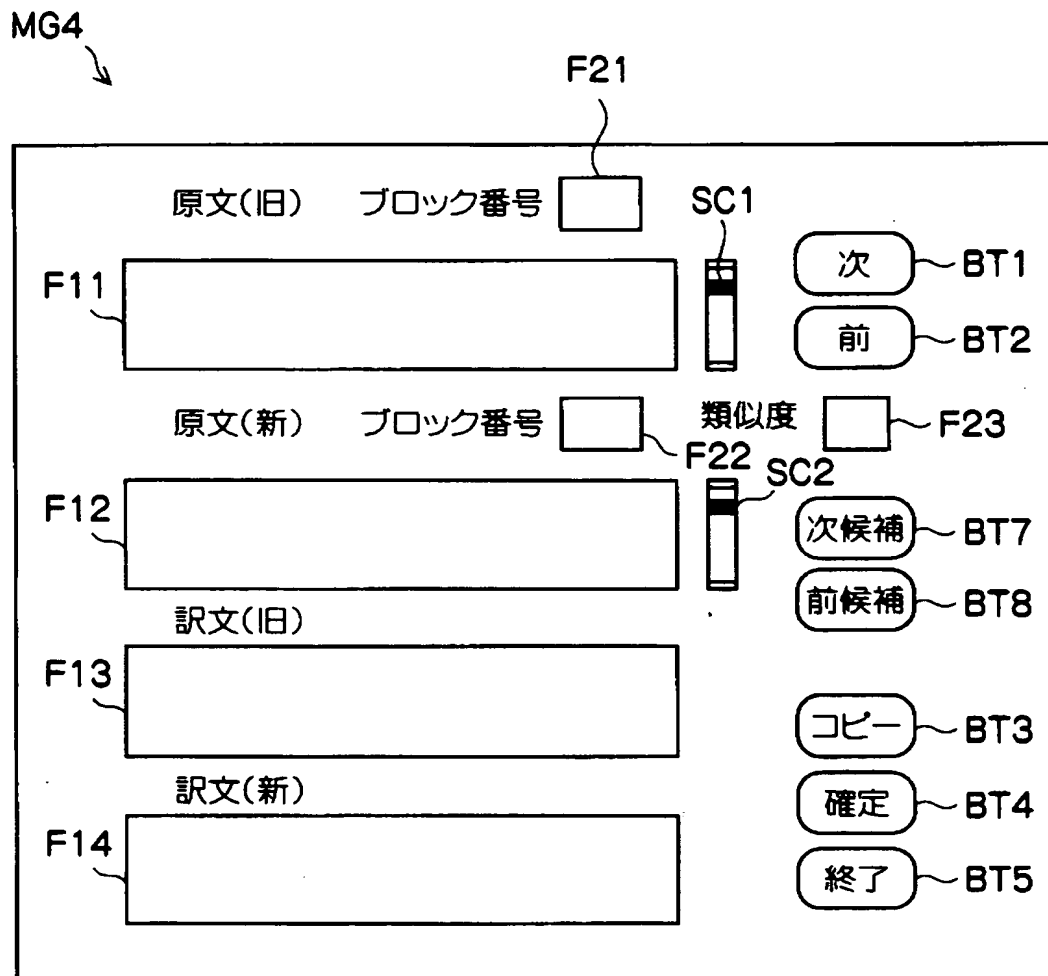
【図 12】



【図 13】



【図 14】



【図 1 5】

文	文番号
文 1。	1
文 2。	2
文 3。	3
文 4。	4
文 5。	5
文 6。	6,18
文 7。	7
文 8。	8
文 9。	9
文 10。	10
文 11。	11
文 12。	12
文 13。	13
文 14。	14
文 15。	15
文 16。	16
文 17。	17

【図 1 6】

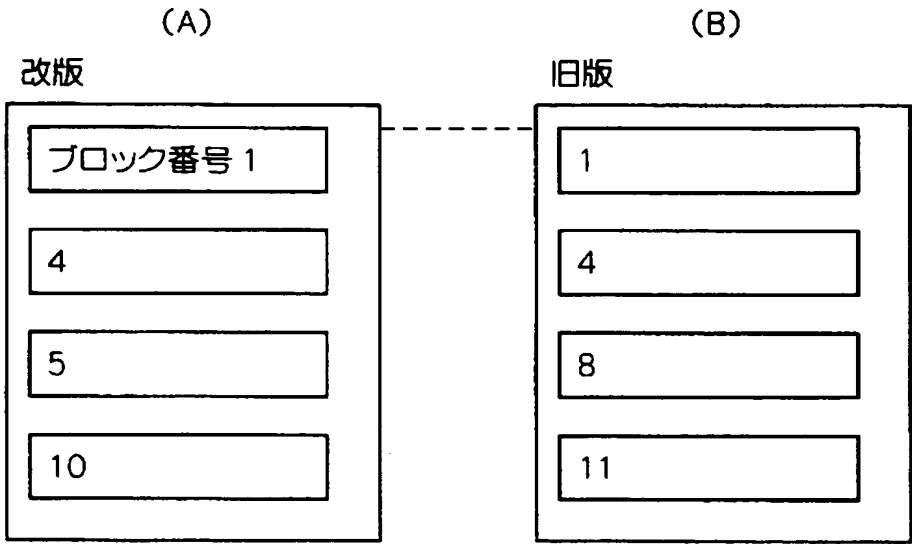
改版文番号	旧版文番号
A	1
B	2
C	3
D	4
E	6
F	7
G	14
H	15
I	12
J	13
K	9
L	10
M	
N	
O	
P	16
Q	17

【図 1 7】

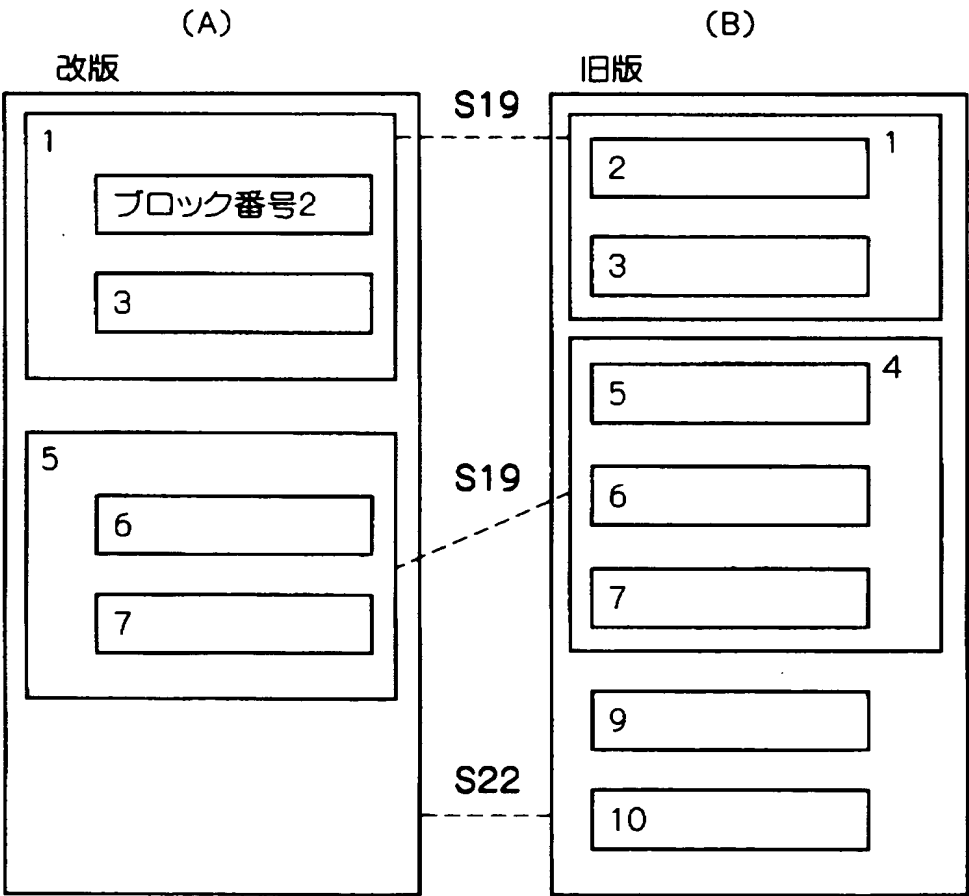
	改版ブロック番号	旧版ブロック番号	類似度
L21	10	8	100
L22	1	1	92
L23	5	4	47
L24	1	11	25
L25	4	4	22
L26	1	4	13

階層位置 $i=1$ の時のブロック組み合わせ表

【図 18】



【図 19】



【書類名】 要約書

【要約】

【課題】 文書と文書の対応関係を、その文脈に配慮して検出する。

【解決手段】 第 1 の文書の構成要素と第 2 の文書の構成要素のあいだの対応関係を検査する文書対応関係検査装置であって、第 1 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析すると共に、第 2 の文書の構成要素のうち、少なくとも 1 つ以上の文を含む文ブロックの論理的な構造を解析する論理構造解析部と、論理構造解析部の解析結果をもとに、第 1 の文書の文ブロックと第 2 の文書の文ブロックのあいだの対応関係を検出する対応関係検出部とを備える。このような構成において、対応関係検出部が、同じ文書の文ブロックのなかに階層構造が存在する場合、上位の階層の文ブロックに関する対応関係を検出したあと、下位の階層の文ブロックの対応関係を検出することも好ましい。

【選択図】 図 1

特願 2 0 0 3 - 1 4 8 6 5 7

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 0 2 9 5]

1. 変更年月日
[変更理由]

1 9 9 0 年 8 月 2 2 日
新規登録

住 所
氏 名

東京都港区虎ノ門1丁目7番12号
沖電気工業株式会社